

Genfore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

us nucleotide - nucleotide search, using sw model

Run on: August 25, 2000, 18:23:19 ; Search time 984.13 Seconds

(without alignments)
1900.641 Million cell updates/sec

Title: US-09-257-585-1

Perfect score: 1047
Sequence: 1 gatctgttaataatcatatat.....atgagagagagagagagatcc 1047

Scoring table: IDENTITY_MMC
Gapop 10.0 ; Gapext 1.0

Searched: 972840 seqs, 89244905 residues

Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database :

- 1: GenBank:*
- 2: gb_pat:*
- 3: gb_baz:*
- 4: gb_om:*
- 5: gb_lov:*
- 6: gb_prl:*
- 7: gb_prl:*
- 8: gb_prl:*
- 9: gb_prl:*
- 10: gb_prl:*
- 11: gb_prl:*
- 12: gb_prl:*
- 13: gb_prl:*
- 14: gb_prl:*
- 15: gb_prl:*
- 16: gb_prl:*
- 17: gb_prl:*
- 18: gb_prl:*
- 19: gb_prl:*
- 20: gb_prl:*
- 21: gb_prl:*
- 22: gb_prl:*
- 23: gb_prl:*
- 24: gb_prl:*
- 25: gb_prl:*
- 26: gb_prl:*
- 27: gb_prl:*
- 28: gb_prl:*
- 29: gb_prl:*
- 30: gb_prl:*
- 31: gb_prl:*
- 32: gb_prl:*
- 33: gb_prl:*
- 34: gb_prl:*
- 35: gb_prl:*
- 36: gb_prl:*
- 37: gb_prl:*
- 38: gb_prl:*
- 39: gb_prl:*
- 40: gb_prl:*
- 41: gb_prl:*
- 42: gb_prl:*
- 43: gb_prl:*

- 44: gb_prl:*
- 45: gb_prl:*
- 46: gb_prl:*
- 47: gb_prl:*
- 48: gb_prl:*
- 49: gb_prl:*
- 50: gb_prl:*
- 51: gb_prl:*
- 52: gb_prl:*
- 53: gb_prl:*
- 54: gb_prl:*
- 55: gb_prl:*
- 56: gb_prl:*
- 57: gb_prl:*
- 58: gb_prl:*
- 59: gb_prl:*
- 60: gb_prl:*
- 61: gb_prl:*
- 62: gb_prl:*
- 63: gb_prl:*
- 64: gb_prl:*
- 65: gb_prl:*
- 66: gb_prl:*
- 67: gb_prl:*
- 68: gb_prl:*
- 69: gb_prl:*
- 70: gb_prl:*
- 71: gb_prl:*
- 72: gb_prl:*
- 73: gb_prl:*
- 74: gb_prl:*
- 75: gb_prl:*
- 76: gb_prl:*
- 77: gb_prl:*
- 78: gb_prl:*
- 79: gb_prl:*
- 80: gb_prl:*
- 81: gb_prl:*
- 82: gb_prl:*

Pred. No. is the number of results predicted by
score greater than or equal to the score of the
and is derived by analysis of the total score.

SUMMARIES

Result No.	Score	Query Match	Length	Hit
1	44.2	4.2	44.2	AC034149
2	44	4.2	45.0	U00000
3	43.4	4.1	47.5	U00000
4	43	4.1	45.4	AB001462
5	42.8	4.1	47.7	U00000
6	42.6	4.1	47.7	AF165926
7	41.8	4.0	47.4	AF165926
8	41.6	4.0	47.2	AF165926
9	40.8	3.9	46.4	AF165926
10	40.8	3.9	46.4	AF165926
11	40.6	3.9	46.4	AF165926
12	40.4	3.9	46.4	AF165926
13	40.4	3.9	46.4	AF165926
14	40	3.8	46.4	AF165926
15	40	3.8	46.4	AF165926
16	39.8	3.8	46.4	AF165926
17	39.6	3.8	46.4	AF165926
18	39.6	3.8	46.4	AF165926
19	39.6	3.8	46.4	AF165926
20	39.6	3.8	46.4	AF165926
21	39.6	3.8	46.4	AF165926
22	39.2	3.8	46.4	AF165926
23	39.2	3.8	46.4	AF165926
24	39.2	3.8	46.4	AF165926

3305	3325:	contig of 51 bp in length
3356	3365:	gap of unknown length
3366	3377:	contig of 512 bp in length
3378	3387:	gap of unknown length
3388	3395:	contig of 512 bp in length
4300	4309:	gap of unknown length
4310	4329:	contig of 520 bp in length
4330	4338:	gap of unknown length
4840	5388:	contig of 519 bp in length
5759	5368:	gap of unknown length
5769	5368:	contig of 520 bp in length
5889	5898:	gap of unknown length
2639	6242:	contig of 344 bp in length
6243	6252:	gap of unknown length
6253	6270:	contig of 518 bp in length
6271	6280:	gap of unknown length
6781	7239:	contig of 479 bp in length
7260	7269:	gap of unknown length
7270	7269:	contig of 512 bp in length
7782	7791:	gap of unknown length
7792	8119:	contig of 328 bp in length
8120	8129:	gap of unknown length
8130	8409:	contig of 280 bp in length
8410	8419:	gap of unknown length
8420	8925:	contig of 496 bp in length
8426	8925:	gap of unknown length
8926	9437:	contig of 512 bp in length
9438	9437:	gap of unknown length
9438	9644:	contig of 187 bp in length
9645	9644:	gap of unknown length
9645	10164:	contig of 520 bp in length
10165	10174:	gap of unknown length
10175	10693:	contig of 519 bp in length
10694	10703:	gap of unknown length
10704	11182:	contig of 479 bp in length
11183	11192:	gap of unknown length
11193	11710:	contig of 518 bp in length
11711	11720:	gap of unknown length
11721	12238:	contig of 518 bp in length
12339	12348:	gap of unknown length
12349	12760:	contig of 512 bp in length
12761	12779:	gap of unknown length
12771	13283:	contig of 513 bp in length
13284	13293:	gap of unknown length
13294	13804:	contig of 511 bp in length
13805	13811:	gap of unknown length
13815	14336:	contig of 512 bp in length
14327	14336:	gap of unknown length
14337	14840:	contig of 504 bp in length
14431	14850:	gap of unknown length
14451	15432:	contig of 512 bp in length
15432	15472:	gap of unknown length
15663	15672:	contig of 519 bp in length
15673	15891:	gap of unknown length
15902	15921:	gap of unknown length
16071	16670:	contig of 769 bp in length
16671	16680:	gap of unknown length
16681	17128:	contig of 448 bp in length
17129	17188:	gap of unknown length
17139	17638:	contig of 520 bp in length
17659	17688:	gap of unknown length
17659	17719:	contig of 51 bp in length
17720	17739:	gap of unknown length
17730	18155:	contig of 416 bp in length
18146	18155:	gap of unknown length
18156	18661:	contig of 506 bp in length
18662	18671:	gap of unknown length
18672	19183:	contig of 512 bp in length
19184	19193:	gap of unknown length
19194	19593:	contig of 400 bp in length
19594	19653:	gap of unknown length
19604	20113:	contig of 510 bp in length
20114	20880:	gap of unknown length
20124	20880:	contig of 757 bp in length

20881	20890	gap of unknown length
20891	21248	contig of 358 bp in length
21248	21258	gap of unknown length
21258	21756	contig of 518 bp in length
21756	21777	gap of unknown length
21777	22184	contig of 368 bp in length
22184	22185	gap of unknown length
22185	22465	contig of 456 bp in length
22465	22471	gap of unknown length
22471	22481	contig of 519 bp in length
22481	23150	gap of unknown length
23150	23160	contig of 351 bp in length
23160	23321	gap of unknown length
23321	24041	contig of 511 bp in length
24041	24042	gap of unknown length
24042	24552	contig of 511 bp in length
24552	25074	contig of 512 bp in length
25074	25085	gap of unknown length
25085	25548	contig of 464 bp in length
25548	25559	gap of unknown length
25559	26070	contig of 512 bp in length
26070	26081	gap of unknown length
26081	26597	contig of 517 bp in length
26597	26608	gap of unknown length
26608	27095	contig of 488 bp in length
27095	27106	gap of unknown length
27106	27325	contig of 519 bp in length
27325	27345	gap of unknown length
27345	28145	contig of 510 bp in length
28145	28475	contig of 521 bp in length
28475	28685	gap of unknown length
28685	29204	contig of 519 bp in length
29204	29495	gap of unknown length
29495	29726	contig of 512 bp in length
29726	29727	gap of unknown length
29727	29745	contig of 512 bp in length
29745	30259	gap of unknown length
30259	30771	contig of 512 bp in length
30771	31281	gap of unknown length
31281	31289	contig of 448 bp in length
31289	31558	gap of unknown length
31558	32286	contig of 519 bp in length
32286	32297	gap of unknown length
32297	32815	contig of 519 bp in length
32815	32826	gap of unknown length
32826	33445	contig of 519 bp in length
33445	33455	gap of unknown length
33455	33882	gap of unknown length
33882	34400	contig of 518 bp in length
34400	34411	gap of unknown length
34411	34906	contig of 496 bp in length
34906	34916	gap of unknown length
34916	35342	contig of 408 bp in length
35342	35345	gap of unknown length
35345	35856	contig of 512 bp in length
35856	36352	contig of 496 bp in length
36352	36453	gap of unknown length
36453	36810	contig of 448 bp in length
36810	36821	gap of unknown length
36821	37345	contig of 515 bp in length
37345	37846	contig of 519 bp in length
37846	37874	gap of unknown length
37874	38245	contig of 520 bp in length
38245	38405	gap of unknown length
38405	38926	contig of 512 bp in length
38926	38917	gap of unknown length

[illegible]

06 411 TATGTCACCATTTTGGCAATTGGGAATTTTTATTTAAGTAACTAAGTAATAATGAACAATAI 372
07 82 attaatcttgcgtttacatcacaatttaaaccaatttcgctccgaagaagaaattttaagttagata 141

[illegible]

Ddb 4 k1 AAAATAGTAA**T**TTT17AA1AIAIAIAAAAAATPAC1AI1IAAT1TA 42⁹

RESULT 4
AP0014826

ACCESSION	AE001482	AE001483
LOCUS	AE001482	AE001483
DEFINITION	Helicobacter pylori, strain J59 section 43 of the complete genome.	Helicobacter pylori, strain J59 section 43 of the complete genome.
DATE	29 JAN 1999	29 JAN 1999

ORGANISM	<i>Helicobacter pylori</i> J99, <i>Helicobacter pylori</i> J99
SOURCE	
KEYWORDS	
VERSION	AF001482.1
RELEASE	APR01482.1
CL	GI:4155018

REFERENCES
Aim, R. A., Ling, D., S. L., Moll, D. W., King, R. L., Brown, R. D.,
1 (bases 1 to 1549)
bacterial, Proteobacterid, epsilon subdivision, Helicobacter q
Helicobacter.
1 (bases 1 to 1549)

Deig, P. C., Smith, D. R., Noonan, B., Gault, B. C., deJonge, E. L., Carnell, G., Tunney, P. J., Caruso, A., O'Lea Nickelsson, M., Mills, T., Ives, C., Gibson, R., Werbberg, D., Mills, S. D., Jiang, Q., Taylor, J., Vovis, C. F. and Trust, T. J.

TITLE Antigenic sequence comparison of two unrelated isolates of the gastric pathogen *Helicobacter pylori*
JOURNAL Nature 397 (6715), 176-180 (1999)
KEYWORDS 99120557

PERMANENTLY PUBLISHED CITATION APPEARS IN NATURE 1999 FEB
 25, 397(6721):7191
 REFERENCE
 2. (bases 1 TO 1549)
 KIM, H. L., ALM, K. A. and TRUST, J. J.
 TITLE
 Insect Subspecies

TITLE Direct Submission
JOURNAL Submitted (12 JAN 1993) Astria Research Center Boston, 128 State Street, Cambridge, MA 02139, USA
COMMENT A. Astria Research Center Boston, 128 State Street, Cambridge, MA 02139, USA
Address all correspondence to: hp.arch@astria.com or Rich.Astria@astria.com

BRADYON G. GULD, Gillies Cancer, Anthony Curuso, Debra M. Miller, Rene Gibson, and Gerald F. Vivas are with Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA, 02453. **Gün Jünger**

Dr. R. Taylor and the University of Alberta Department of Microbiology and Immunology, Edmonton, Alberta, Canada, 297 and the Canadian Bacterial Diseases Network. All other authors with Astra Research Center Boston, 120 Spring Street,

Cambridge, MA, 02139. Putative identifications, sequence alignments, and name and sequence search capability are available at ACRB's World Wide Web site. (URL: <http://www.astra-boston.com/hpy1011>).

FEATURES	SOURCE	Location/Qualifiers
1.	15549	/organism="Helicobacter pylori J99"
		/strain="J99"

```

qctno      /db_xref="taxon:85963"
           222,.878
/qcne="Orf18"
/nuc0e="Jhp0480"
222,.878

```

CD5
2.2.2.878
/gene="COL18"
/note="Similar to H. pylori 26695 gene HPO511"
/codon_start=1
/transl_table=1

```

/protein_id "A006068.1"
/product "cag isiaid protein"
/protein_id "A006068.1"
/ab_xref "01:155029"
/definition "MNTFTHSHND NAEISNSNSHII NININISYKCS"

```

HEALTHCARE • **RESEARCH** • **TECHNOLOGY** • **PHARMACEUTICALS**

PAGE COUNT 47095 a 35017 c 34525 q 48139 t 21 others
ORIGIN

Query Match	4.00E+000	Score: 41.87	E-Inf: 7.76
Best Local Similarity	55.00%	Ident. No.: 5/4	47.7%
Matches: 82;	Conserved: 82;	Mismatches: 0	
QY 52	cttttaaaatttlaacgatttataatctgaaatcaatttgcattt		
Db 77374	CTGCTTAACTTTTCAAAATTAATTAATTAATTAATCTTTT		
QY 112	attcgttctcaagagatttttatttcaagatttgcattttttta		
Db 77434	ATCTCTTAAGGACATAATTAATTAATTAATTAATCTATAT		
QY 172	aaatacatgctctttaaataatcatttttgcctg		200
Db 77494	ATTATATATTTTTTTTAAATATATATATATATATATATAT		77552

```

Query Match Similarity 4.1% Score 42.67 DB 39; Length 100000
Best Local Similarity 41.1% Prod No. 3.4; Mismatches 44; Indels 0; Gaps 0;
Matches 69; Conservative 0;

UY      88   ttatgaattctacagaataatgaactgtcgaaaaagtgatttaaacdaaatlqqc 147
            ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    36786 ttttaatctgcacacattttaccattttccaaattttccaaatggcatatataaaccccaaaatatat 36845
            ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

UY      148 atttllttaacatcatttttaaanaaatatcatcgtcttaaacgacctttcca 200
            ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    36846 atattatttatcatttttttttatsaacctaaactgatcattatccagcatctcna 36894
            ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
AB026647 A1542 bp DNA PLN 07 MAY-1999
LOCUS     Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MDL12,
DEFINITION complete sequence.
ACCESSION AB026647
VERSION   AB026647.1 GI:4757403
KEYWORDS  Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
SOURCE    Arabidopsis thaliana
ORGANISM  Arabidopsis thaliana
           Eukaryota; Viridiplantae; streptophyta; Embryophyta; tracheophyta;
           eudicotyledons; Spermatophytes; Magnoliopsida; eudicotyledons; core
           eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae;
           Arabidopsis.
REFERENCE 1 (sites)
AUTHORS   Nakamura,K.
TITLE     Structural Analysis of Arabidopsis thaliana Chromosome 3. 11
JOURNAL   Unpublished (1999)
REFERENCE 2 (bases 1 to 81542)
AUTHORS   Nakamura,Y.
TITLE     Direct Submission
JOURNAL   Submitted (28-Nov-1999) to the dbj/psm/j/v/genbank databases.
          Yabuuchi,Nakamura, Koyama DNA Research Institute, Laboratory of
          Gene Structure 2; Yada 1532-3, Misaki-daigaku, Chiba 245-0285, Japan
          (E-mail:yabukamu@kayama.or.jp, Tel:81-438-52-3935(fax:7443).
          Fax:81-438-52-3934)
FEATURES
     source             Location/Junkifiers
                        1.. 81542
                        /organism="Arabidopsis thaliana"
                        /strain="Columbia"
                        /db_xref="taxon:3702"
                        /chromosome="3"
                        /clone="MDL12"
                        /clone-lib="Mitsui P1"
HASH COUNT
25859 a 14006 c 26843 g

```

[illegible]

[illegible]

	/stratn-"SI"
	/db_xref-"taxon:5908"
	/clone-"HindIII/Sau3A"
rRNA	670. .2984
	/gene-"LSU-beta"
	/product-"large subunit ribosomal RNA"
gene	670. .2984
	/gene-"LSU-beta"
rRNA	2987. .3067
	/gene-"rRNA-Leu"
	/product-"Transfer RNA-Leu"
gene	2987. .3067
	/gene-"rRNA-Leu"
rRNA	3078. .3357
	/gene-"LSU-alpha"
	/product-"large subunit ribosomal RNA"
gene	3078. .3357
	/gene-"LSU-alpha"
rRNA	3373. .3439
	/partId1
	/gene-"rRNA-Met"
	/product-"transfer RNA-Met"
gene	3373. .3439
	/gene-"tRNA-met"
BASE COUNT	1337 A 483 G 1242 C
STRUCTURE	

Query Match	Similarity	Score	DB	Length
Post Local	52.4%	Prot. No. 17		
Matches	88	Conservative	6	Mismatches
			80	Indels
			0	Gaps

Query	Subject	Score	DB	Length
1	5	54	GenBank	149
2	11	51	GenBank	149
3	12	51	GenBank	149
4	13	51	GenBank	149
5	14	51	GenBank	149
6	15	51	GenBank	149
7	16	51	GenBank	149
8	17	51	GenBank	149
9	18	51	GenBank	149
10	19	51	GenBank	149
11	20	51	GenBank	149
12	21	51	GenBank	149
13	22	51	GenBank	149
14	23	51	GenBank	149
15	24	51	GenBank	149
16	25	51	GenBank	149
17	26	51	GenBank	149
18	27	51	GenBank	149
19	28	51	GenBank	149
20	29	51	GenBank	149
21	30	51	GenBank	149
22	31	51	GenBank	149
23	32	51	GenBank	149
24	33	51	GenBank	149
25	34	51	GenBank	149
26	35	51	GenBank	149
27	36	51	GenBank	149
28	37	51	GenBank	149
29	38	51	GenBank	149
30	39	51	GenBank	149
31	40	51	GenBank	149
32	41	51	GenBank	149
33	42	51	GenBank	149
34	43	51	GenBank	149
35	44	51	GenBank	149
36	45	51	GenBank	149
37	46	51	GenBank	149
38	47	51	GenBank	149
39	48	51	GenBank	149
40	49	51	GenBank	149
41	50	51	GenBank	149
42	51	51	GenBank	149
43	52	51	GenBank	149
44	53	51	GenBank	149
45	54	51	GenBank	149
46	55	51	GenBank	149
47	56	51	GenBank	149
48	57	51	GenBank	149
49	58	51	GenBank	149
50	59	51	GenBank	149
51	60	51	GenBank	149
52	61	51	GenBank	149
53	62	51	GenBank	149
54	63	51	GenBank	149
55	64	51	GenBank	149
56	65	51	GenBank	149
57	66	51	GenBank	149
58	67	51	GenBank	149
59	68	51	GenBank	149
60	69	51	GenBank	149
61	70	51	GenBank	149
62	71	51	GenBank	149
63	72	51	GenBank	149
64	73	51	GenBank	149
65	74	51	GenBank	149
66	75	51	GenBank	149
67	76	51	GenBank	149
68	77	51	GenBank	149
69	78	51	GenBank	149
70	79	51	GenBank	149
71	80	51	GenBank	149
72	81	51	GenBank	149
73	82	51	GenBank	149
74	83	51	GenBank	149
75	84	51	GenBank	149
76	85	51	GenBank	149
77	86	51	GenBank	149
78	87	51	GenBank	149
79	88	51	GenBank	149
80	89			


```

gene
/db_xref="GI:5306133"
/translation="MTRKLR...RNSLQV...MHWV...IN...PAI...ST...LY...LV...FY...
PLNENLYYCTIMK...EL...GNEV...G...K...R...J...N...M...R...S...M...R...P...T...I...G...R...S...Q...W...
V...L...N...Y...S...S...P...R...K...V...Y...A...V...S...V...F...R...L...F...F...E...P...R...I...E...K...S...R...L...I...N...F...R...R..."
complement(7574..7870)
/feature="rps19"
complement(7574..7870)
/gene="rps19"
/codon_start=1
/transl_table=4
/product="ribosomal protein S19"
/protein_id="A0A01918.1"
/db_xref="GI:5306134"
/translation="MTRK...M...K...S...M...A...T...K...S...M...S...N...V...A...N...K...A...N...K...L...S...I...M...N...S...S...T...P...I...
S...I...E...N...T...E...F...E...P...R...K...V...Y...A...V...S...V...F...R...L...F...F...E...P...R...I...E...K...S...R...L...I...N...F...R...R..."
complement(7879..8667)
/gene="rpl2"
complement(7879..8667)
/feature="rpl2"
complement(7879..8667)
/gene="rpl2"
/codon_start=1
/transl_table=4
/product="ribosomal protein L2"
/protein_id="A0A01919.1"
/db_xref="GI:5306135"
/translation="MTRK...M...K...S...M...A...T...K...S...M...S...N...V...A...N...K...A...N...K...L...S...I...M...N...S...S...T...P...I...
S...I...E...N...T...E...F...E...P...R...K...V...Y...A...V...S...V...F...R...L...F...F...E...P...R...I...E...K...S...R...L...I...N...F...R...R..."
complement(8668..9456)
/gene="rpl2"
complement(8668..9456)

```

Query Match 4.88; Score 40; DB 34; Length 47296;
 Best Local Similarity 52.44; Prod. No. 15;
 Matches 88; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

UY 5 tgaataacatataatatttgcgcatttgcgaacgcctattctttaaatc 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
UB 43856 TGTATTTTATTATTAATTTTATATTTTATTTGATTTGATTTGATTAATTAATTT 43915
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
UY 65 taacgaattgaagacgaatctaatcttcaagctctatgaagaatcaatccgaag 124
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
UB 43916 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 43975
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
UY 125 aatgaatgaagcaaatgaacatttttaacgaatlllaaaga 172
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
UB 43976 AAGCTTTTAACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 44023

```

Search completed: August 25, 2000, 22:56:56
 Job time: 16388 sec

F1	Dorjann W., Chen G., Meyermans II., Van Montagu M.
F2	WPI; 99-190169/716.
P1	Poplar cinnicoyl-CoA-o-methyltransferase (CcoAMT) promoter - to
P2	provide transgenic plants having altered lignin level and/or
P3	composition
P5	Claim 1: File ID: 69pp; English.
CC	This sequence represents the Poplar (<i>Populus trichocarpa</i>)
CC	cinnicoyl-CoA-o-methyltransferase (CcoAMT) promoter, PTCOAMT1, of the
CC	invention. The promoter is specific for plant of tree living vessel
CC	elements and/or xylem adjacent ray cells. The promoter is used in genetic
CC	modification to provide transgenic plants having altered lignin level
CC	and/or composition, which improves digestibility of forages and pulping
CC	properties of trees.
SQ	Sequence 3800 BP; 1225 A; 668 C; 791 G; 1206 T;
Query Match:	5.4%; Score 36; DB 1; Length 3800;
Hect. Local Similarity:	49.5%; Prev. No. 0.73;
Matches 95; Conservative 0; Mismatches 95; Indels 0; Gaps 0	
Q7	42 accgcctatcctcttttaaatctttagcgattagagtcgacatcaatttcgtctaca tct
Db	1 1
	800 ATCGACATGATCTTTATAAAAAATAAAAAAATATATTATTAACCATTTCCAGIAAAAAA 741
Q7	102 caattaaagaattcgtccaaaaggatllaaqqaagtllaqcatlttllaatera tct
Db	1 1
	740 CACTTGAAAAGCACCATTAACCATTTTTCAGTGAACATTAATAAAATTTTCACA 681
Q7	162 ttltlaaaaaataaacatcgcgtttaaccacaccttgacgaagatcacgcgtacacagc 221
Db	1 1
	680 TGTTTCCAAAATATATATACAAACATTGTCAAACTGGANAGTAGTATTTCATTATATTT 621
Q7	222 ctggatga cgg
Db	1 1
	620 TTCTTTA 613
RESULT 9	
ID	V74608/c
ID	V74608 standard; DNA; 3055 BP.
AC	V74608;
DE	16-MAR-1999 (first entry)
KW	Staphylococcus aureus conflg SRO ID #297.
KW	Computer readable medium; vaccine; Stauerus infection; immunorection;
KW	cellulites; cycloid infection; food poisoning; osteomyelitis; therapy;
KW	skin infection, surgical wound infection, scaled skin syndrome;
KW	toxic shock syndrome; ds.
KW	Staphylococcus aureus.
PH	Key
FT	misc_feature
FT	841..900
FT	/tag= a
FT	/note= "these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
FT	
FT	misc_feature
FT	241..270
FT	/tag= b
FT	/note= "these bases represent a line of missing text in
FT	the sequence listing in the specification. They
FT	are included to maintain the nucleotide numbering
FT	given in the specification for this DNA sequence"
PN	EP-786519-A2.
PD	30-JUL-1997.
PP	07-JAN-1997; 100117
PR	05-JAN-1996; US-009861.
PA	(HUKA-) HUMAN GENOME SCI INC.
P1	Barrash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
P1	Rosen CA;
DR	WPI; 97-324923/75.
PT	polynucleotides) and proteins derived from Staphylococcus aureus
PT	stored on computer readable medium and used in the production of
PT	anti-S.aureus vaccines
CS	claim 1; Page 1123-1125; 3271pp; English.

STATE: NEW YORK
COUNTRY: USA
ZIP: 10151
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1 0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 05/06/816,1558
FILING DATE: 12-MAR-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: KOMALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454310-2490
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 14877 base pairs
TYPE: nucleic acid

STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-816-155B-8

Query Match 3.18, Score 32.8, DB 4, Length 19877;
 Best Local Similarity 56.58; Pred. No. 4.5;
 Matches 61; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 15 atatatatttaccgcatlltgcataccgcatallctllttaaatttaccgcatllaa 74
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 2148 AIAAATATTTTGAAGATGATGATGATGATGATGATGATGATGATGATGAT 2207
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 75 gaacgcatatttgaatttgaatttgaatttgaatttgaatttgaatttgaatttga 122
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 2208 GAAACATATATGACATATATATATATATATATATATATATATATATATATAT 2255
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4

US-08-451-405A-2/c
 Sequence 2, Application US/08/451405A
 Patent No. 5736458
 GENERAL INFORMATION:
 APPLICANT: FASSEL, NICOLAS JOSEPH
 TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
 TITLE OF INVENTION: METHOD FOR EXPRESSING A DICTYOSTELID PROTEIN
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESS: THE WEBB LAW FIRM
 STREET: 700 KOPPERS BUILDING, 430 SEVENTH AVENUE
 CITY: PITTSBURGH
 STATE: PENNSYLVANIA
 COUNTRY: UNITED STATES OF AMERICA
 ZIP: 15219-1818
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" FLOPPY DISK
 COMPUTER: MIDWEST MICRO 486-50
 OPERATING SYSTEM: DOS
 SOFTWARE: WORDPERFECT 6.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/451,405A
 FILING DATE: 26-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/965,273
 FILING DATE: 15-JAN-1994
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 741
 TYPE: NUCLEIC ACID
 STRANDEDNESS: SINGLE
 TOPOLOGY: UNKNOWN
 US-08-451-405A-2

Query Match 3.18, Score 32; DB 2; Length 731;
 Best Local Similarity 51.08; Pred. No. 1.7;
 Matches 102; Conservative 0; Mismatches 95; Indels 3; Gaps 1;

QY 2 atcttataatataatatttatttgcggaatttgcgaacctatattttttaaana 61
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 252 ATCCAAAATATCTGAAATTTTITTAGAATTTCTTA--TCAATATACCTGACAAA 196
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 62 tttaataatataatataatatttgcggaatttgcgaacctatattttttaaana 121
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 195 TCAATTTTACGATCATCATCATATATATATATATATATATATATATATATATAT 136
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 122 aatgaatatttgaatgaatatttgcggaatttgcgaacctatattttttaaana 181
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 145 TAGAAATATTTATGTAATATATATATATATATATATATATATATATATATATAT 76
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 182 cttaccacacatttgcgaacctatattttttaaana 291
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 75 GAAAAAAT 34

RESULT 5

US-08-126-593A-1
 Sequence 1, Application US/08/126593A
 Patent No. 5527700
 GENERAL INFORMATION:
 APPLICANT: Kaslow, David J.
 TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
 TITLE OF INVENTION: METHOD FOR EXPRESSING A DICTYOSTELID PROTEIN
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESS: TANGHEE AND TANGHEE AND TANGHEE AND TANGHEE
 STREET: ONE MARKET PLACE, STOWART STREET, LOWEY
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94105-1492
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: MS-DOS/MS-WINDOWS
 SOFTWARE: Patent Release #1.0, Version #1.00
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/126,593A
 FILING DATE: 22-MAY-1994
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/01,294
 FILING DATE: 10-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Hashtab, Kevin L.
 REGISTRATION NUMBER: 34,774
 REFERENCE/DOCKET NUMBER: 15,281-46-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 943-9600
 TELEFAX: (415) 943-9604
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 858 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1-858
 OTHER INFORMATION: /product "Pgs28 surface" for
 OTHER INFORMATION: from Plasmodium gallinaceum
 US-08-126-593A-1

Query Match 3.74, Score 11.4; DB 1;
 Best Local Similarity 52.79; Pred. No. 2.8;
 Matches 68; Conservative 0; Mismatches 61;

QY 50 tctctttaaatttgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 114
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 2 TTTTGCAAT 274
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 110 caatttcgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 114
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 62 CAAAT 222
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 170 aaaaataat 178
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 122 AATGAAAT 140
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Query Match	4.0%	Score 31.4	DB 2	Length 858
Best Local Similarity	52.7%	Prod. No. 4.8		
Matches	68	Conservative	0	Mismatches 61
				Indels 0
				Gaps 0
CY	50	ttctttttaaaatttcaacgatttaaaacgcgaattcaatttgcgactacagaattaa	109	
1b	1		1	
1b	2	tttttttcattattattatcatatttttaaaattcaattttatcttccattaa aaattattccaa	61	
CY	110	caatttcgtccaaaagaagatatttcttgaagatcttgcgtatcttgcgtatcttgaat	150	
1b	1		1	
1b	62	caaaatatttcaaacgaattatatttacc aaacgaacaaacgaattttatattttattt aaaaa	121	
CY	170	aaaattacat	178	
1b	1		1	
1b	122	aattgaataat	130	

```

US-08-319-704-5/c
? Sequence 5: Application US/08319704
? Patent No. 581617
? GENERAL INFORMATION:
? APPLICANT: Hollman, Stephen L.
? ATTORNEY: Charocovitz, Yipin
? APPLICANT: Hedstrom, Richard C.
? APPLICANT: Iwolan, Denise L.
? TITLE OF INVENTION: Protective 17 kDa Malaria hepatic and
? TITLE OF INVENTION: Erythrocytic Stage Immunogen and Gene
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESS: Naval Medical R & D Command
? STREET: Bldg 1, T-12, 8901 Wisconsin Avenue
? CITY: Bethesda
? STATE: Maryland
? COUNTRY: U.S.A
? ZIP: 20884-5606
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/319,704
? FILING DATE: 07-OCT-1994
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: A. David Spevack
? REGISTRATION NUMBER: 24,743
? REFERENCE/JOCKET NUMBER: 75,206
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (301) 295-6759
? TELEFAX: (301) 295-1022
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1422 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
US-08-319-704-5

Query Match 3.0%; Score 31.2; E-2; Length 1422;
Best Local Similarity 48.8%; Pred. No. 4.1;
Matches 84, Conservative 0, Mismatches 88, Indels 0, Gaps 0.

0Y 7 aaattacatatcttattttgttcgcagatttgcatacagcgcattctttaaattta 56
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 457 TAAATAATAAATAATATATCTTTGAGAGAAATCTTAATATATGTTTATATGATTTTAA 298
6Y 67 acagattaaagacgaatccaattttagcgtctacacgaattaaacgaatcgcgaagaag 126
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 TTGCAATTAATAATAAATAAATTTTAATTAATTAATTAATGAAATTAATAATTTTAAATAATTT 218
2Y 127 gtaattagagagagatttgcattttttttaaaccatttttaaaaaaatatcat 178
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 AAATTGAGAAATAATTAATAATATTTTCATACAAAATTAATAATAAATAAT 186

RESULT 8
US-08-062-472B-2
? Sequence 2: Application US/08062472B
? Patent No. 5695954
? GENERAL INFORMATION:
? APPLICANT: Sherwood, Nancy G M
? APPLICANT: Parker, David B
? APPLICANT: McRoary, John E
? APPLICANT: Lesheld, David W
? TITLE OF INVENTION: DNA ENCODING TWO FISH NEUROPEPTIDES
? NUMBER OF SEQUENCES: 49
? CORRESPONDENCE ADDRESS:

```

RESULT 10
 US-08-832-883-48/c
 Sequence 48, Application: 22/000,000
 Patent No. 5807681
 GENERAL INFORMATION:
 APPLICANT: Giordano, Antonio
 APPLICANT: Baldi, Alphonse
 TITLE OF INVENTION: METHOD FOR THE DIAGNOSIS AND
 NUMBER OF SEQUENCES: 115
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEIDEL, GINA, LAV ROMA & MONACO, P.
 STREET: Suite 1800 Two Penn Center Plaza
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19102
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: pc-tos/ms dos
 SOFTWARE: patent in release #1.0, version #1.00
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: 22/000,000
 FILING DATE: 02/06/92 084
 CLASSIFICATION: 445
 ATTORNEY/AGENT INFORMATION:
 NAME: Monaco, Daniel A.
 REGISTRATION NUMBER: 60,46

1

Patent No. 5885589
GENERAL INFORMATION:
APPLICANT: FOGED, Niels L.
APPLICANT: PETERSEN, Svend
TITLE OF INVENTION: PASTORRELLA VACCINE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/453,141
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/293,314
FILING DATE: 22-AUG-1994
APPLICATION NUMBER: US 07/582,945
FILING DATE: 12-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/00084
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 1995/88
FILING DATE: 04-DEC-1988
ATTORNEY/AGENT INFORMATION:
NAME: HENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/WORKET NUMBER: 4539,100,AFCD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4380 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 219..4076
OS-08-453-141-1

Query Match 2.98; Score 30.4; DB 3; Length 4380;
Best Local Similarity 50.78; Prod. No. 12;
Matches 73; Conservative 0; Mismatches 71; Indels 0, Gaps 0;
CY 2 atctggaattatattatatttgcgacttctgtacacgactatctctttaa 61
DB 4191 ATGGTAATACATTTCATTTATCATTCATTCATTCATTCATTCATTAATAATT 4132
CY 62 ttttaacgatttaaaagaaattcaattttaggtttctacacgatttaacatttgcga 121
DB 4131 CTTAAACAATAAATTAAAGCTTATTAATTAATTAAGGCTTAAATACAGTTTAAATTACTTAA 4072
CY 122 aagagattttaagcgaagatttg 145
DB 4071 GTGCTCTGTGTTAAGCGAGGCTTTG 4048

Search completed: August 25, 2000, 22:53:18
Job time: 12255 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

MM nucleic - nucleic search, using SW model

Run on: August 25, 2000, 17:41:26 : Search time 579.47 Seconds

(without alignments)
7969,156 Million cell updates/sec

Title: US-09-257-585-1

Perfect score: 1047

Sequence: 1 gattctgataataataataat.....atgaagagacacacataacc 1047

Scoring table: IDENTITY, NUC

Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: em_est1:*

2: em_est2:*

3: em_est3:*

4: em_est4:*

5: em_est5:*

6: em_est6:*

7: em_est7:*

8: em_est8:*

9: em_est9:*

10: em_est10:*

11: em_est11:*

12: em_est12:*

13: em_est13:*

14: em_est14:*

15: em_est15:*

16: em_est16:*

17: em_est17:*

18: em_est18:*

19: em_est19:*

20: em_est20:*

21: em_est21:*

22: em_est22:*

23: em_est23:*

24: em_est24:*

25: em_est25:*

26: em_est26:*

27: em_est27:*

28: em_est28:*

29: em_est29:*

30: em_est30:*

31: em_est31:*

32: em_est32:*

33: em_est33:*

34: em_est34:*

35: em_est35:*

36: em_est36:*

37: em_est37:*

38: em_est38:*

39: em_est39:*

40: em_est40:*

41: em_est41:*

42: em_est42:*

43: em_est43:*

44: em_est44:*

44: qb_est27:*

45: qb_est28:*

46: qb_est29:*

47: qb_est30:*

48: qb_est31:*

49: qb_est32:*

50: qb_est33:*

51: qb_est34:*

52: qb_est35:*

53: qb_est36:*

54: qb_est37:*

55: qb_est38:*

56: qb_est39:*

57: qb_est40:*

58: qb_est41:*

59: qb_est42:*

60: qb_est43:*

61: qb_est44:*

62: qb_est45:*

63: qb_est46:*

64: qb_est47:*

65: qb_est48:*

66: qb_est49:*

67: qb_est50:*

68: qb_est51:*

69: qb_est52:*

70: qb_est53:*

71: qb_est54:*

72: qb_est55:*

73: qb_est56:*

74: qb_est57:*

75: qb_est58:*

76: qb_est59:*

77: qb_est60:*

78: qb_est61:*

79: qb_est62:*

80: qb_est63:*

81: qb_est64:*

82: qb_est65:*

83: qb_est66:*

84: qb_est67:*

85: qb_est68:*

86: qb_est69:*

87: qb_est70:*

88: qb_est71:*

89: qb_est72:*

90: qb_est73:*

91: qb_est74:*

92: qb_est75:*

93: qb_est76:*

94: qb_est77:*

95: qb_est78:*

96: qb_est79:*

97: qb_est80:*

98: qb_est81:*

99: qb_est82:*

100: qb_est83:*

101: qb_est84:*

102: qb_est85:*

103: qb_est86:*

104: qb_est87:*

105: qb_est88:*

106: qb_est89:*

107: qb_est90:*

108: qb_est91:*

109: qb_est92:*

110: qb_est93:*

111: qb_est94:*

112: qb_est95:*

113: qb_est96:*

114: qb_est97:*

115: qb_est98:*

116: qb_est99:*

```

117: qb_qss13: *
118: qb_qss14: *
119: qb_qss15: *
120: qb_qss16: *
121: qb_qss17: *
122: qb_qss18: *
123: qb_qss19: *
124: em_qss13: *

```

[illegible]

1	54.6	5.1	1101	122	CNS00366
2	46.6	4.5	865	125	CNS00919
3	44.6	4.2	1101	122	CNS00919
4	43.4	4.1	913	123	CNS01572
5	43.2	4.1	1201	123	CNS01611
6	41.6	4.0	1101	122	CNS00112
7	41.4	4.0	817	122	CNS0089K
8	41.4	4.0	1201	122	CNS00763
9	40.4	3.9	286	81	C92267
10	40.4	3.9	511	81	C25641
11	40.4	3.9	1101	122	CNS00919
12	40.2	3.8	411	93	A0008026
13	40.2	3.8	878	123	CNS01611
14	40.2	3.8	1225	123	CNS01611
15	40	3.8	1101	122	CNS0011M
16	39.8	3.8	592	48	A0043294
17	39.8	3.8	1101	122	CNS00241
18	39.6	3.8	1101	122	CNS0038D
19	39.6	3.8	1201	123	CNS0166P
20	39.2	3.7	835	122	CNS009K12
21	39.2	3.7	1101	122	CNS0091MC
22	39.2	3.7	1201	125	CNS0016K
23	39	3.7	192	42	A10220C4
24	39	3.7	441	96	A0403034
25	39	3.7	1069	127	CNS0107G
26	39	3.7	1101	122	CNS0106K
27	38.8	3.7	929	122	CNS003701
28	38.6	3.7	126	25	AA164167
29	38.4	3.7	286	81	C04329
30	38.4	3.7	668	40	A1484706
31	38.4	3.7	1101	122	CNS0000J
32	38.2	3.6	532	42	A107867
33	38.2	3.6	403	42	AA918084
34	38.2	3.6	518	22	AA863354
35	38.2	3.6	126	69	AA235048
36	38.2	3.6	660	119	A2053289
37	38	3.6	884	122	CNS0077Y
38	38	3.6	970	123	CNS0182E
39	37.8	3.6	240	48	A0074262
40	37.8	3.6	132	27	AA483090
41	37.8	3.6	445	88	R59083
42	37.8	3.6	475	27	AA494732
43	37.8	3.6	591	68	R51269
44	37.8	3.6	591	64	AM162119
45	37.8	3.6	932	123	CNS0777Y
					AL064921 Drosophila
					AL073300 Drosophila
					AL074411 Drosophila
					AL103489 Drosophila
					AL106627 Drosophila
					AL078714 Drosophila
					AL071066 Drosophila
					AL067650 Drosophila
					C92687 C92687 D147
					C25641 C25641 D147
					A0097442 Drosophila
					A008935 CIT HSP-2
					AL108793 Drosophila
					AL064771 Drosophila
					AL068473 Drosophila
					A005297 A005297
					A0062443 Drosophila
					AL066491 Drosophila
					AL106687 Drosophila
					A007333 Drosophila
					AL070972 Drosophila
					AL1006209 Drosophila
					A1022064 C179636 X
					A0304034 HS 3229 X
					AL098614 Drosophila
					AL094895 Drosophila
					AL066548 Drosophila
					AA364167 EST147220
					F034329 HSC1X622 X
					A1484706 EST242967
					AL062049 Drosophila
					A107867 C1064204 X
					AA918084 C181406 X
					AA927354 C974310 X
					AA235048 C018406 X
					A0052389 RPT1-24
					AL067426 Dros

RESULTS

LOCUS	LENGTH	DATE
CNS00196	1101 bp	03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR09K10 of RP11-98 library from Drosophila melanogaster (fruit fly)	

Accession	AL063921
Version	AL063921.1
Keywords	GI:4941778
Source	CS.
Organism	Fruit fly, <i>Drosophila melanogaster</i>

1 (bases 1 to 1101)

Direct Submission
Submitted (02-JUN-1995) Genoscope - Centre National de Sequencage

- web : www.genoscope.cns.fr

COMMENT

Identification of this BAC and sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.flycity.org> The BAC Drosophila melanogaster BAC library was prepared by Kazuo Y. Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named R621-36 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the Drosophila strain y2; cn bw sp. The same strain used for the BDGP's 11 and R61 libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or titers for hybridization from the BDGP resource center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

	64	507 others
BRITAIN	201	202
EAST INDIA	64	507 others
FRANCE	201	202
GERMANY	64	507 others
HOLLAND	201	202
ITALY	64	507 others
JAPAN	201	202
NETHERLANDS	64	507 others
RUSSIA	201	202
SPAIN	64	507 others
UNITED STATES	201	202
USSR	64	507 others
WEST GERMANY	201	202
OTHER COUNTRIES	64	507 others

Query Match 5.18; Score 53.8; DB 122; Length 1101;
Best Local Similarity 15.48; Pred. No. 0.00021;
Matches 88; Conservative 248; Mismatches 237; Indels 0;
Gaps 0

36. **gl'aacllq' l'q'aytal'laa' qaaq'q'la'cl'la'cl'q'q' ayyq'aaq'tlat.** 440

630 MHWTTAAATYYTSMYYUHMHHNAHAADWWLTHTWTHAYHWALYHYYYMYCA 689

[illegible]

51 ccccaattgaaagagcccaatctgaaatctcagcaacccctatagccctc 620

[illegible]

810 MOLYBDENYLACETIC ACID 869

Dd 870 WTWATTCWACMTMTHHHMMHMHNNHMCJAHHTTMCMSJJHHHCST'HHNNTMYNHCT'HN 929

[illegible][illegible]

[illegible]

SOURCE

ORGANISM

fruit fly,
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Eurytomera,
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 1101)

AUTHORS

Genoscope,

TITLE

Direct Submission

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrel.genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC and sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Osoegawa and
Aaron Mumberg in Peter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named Bp1-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain yw; cu bw sp; the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BAC Resource Center can be
found at http://bioresources.rockefeller.edu/drosophila_bac.htm.

FEATURES

SOURCE

1..1101
/organism="Drosophila melanogaster"
/db_xref="taxdb:7227"
/clone_id="Bp1-98"
/clone="BACR0118"
/note="end : TET"

BASE COUNT

487 a 157 c 27 g 186 t 344 others

ORIGIN

Query Match

3.88; Score 40; DB 122; Length 1101;

Best Local Similarity 25.78; Pred. No. 14;

Matches 118; Conservative 119; Mismatches 219; Indels 4; Gaps 1;

```

07 158 tcaattttaaataatgacatctcgtttaaaccaccttgcaagagatccgcgtatga 217
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 546 tccaaacimcyayccaaahcncyciaynhccyhaayaaahmymwmcacactmvaaym 595
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
09 218 tggcctgggtggaaaggggggggggggggggggggggggggggggggggggggg 277
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 596 nactgtmgtmiamnmymaanaaanaaiaaamcymhiacaaahatytcttcmwta 655
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
07 278 caactaaagacagacagacagacagacacacacacacacacacacacacacacac 347
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 656 yatvatamktmialaiaiaimtaimtaimtaimtaimtaimtaimtaimtaimtaim 715
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
07 348 gacaaagacagacacacacacacacacacacacacacacacacacacacacacac 397
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 716 tmmamacaiaacaaahaaahimymimmmymaahamimtaimtaimtaimtaimta 775
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
07 798 ttaataataacaaagacacacacacacacacacacacacacacacacacacacac 847
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 776 aaahatitvtayaaahimtaimtaimtaimtaimtaimtaimtaimtaimtaimta 835
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
07 848 tccctgacacacacacacacacacacacacacacacacacacacacacacacac 895
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 846 tcyymnmmttctyymvamtmtmtmtmtmtmtmtmtmtmtmtmtmtmtmtmtmtmt 895
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
07 514 aacatttacttcaaaatattactgttgcgtactgaacacacacacacacacacacac 573
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 896 ctcaatmtatagctgttmttmttmttmttmttmttmttmttmttmttmttmttmt 955
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
07 574 agccacacacacacacacacacacacacacacacacacacacacacacacacacac 613
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
10 956 ttatmymnmimmmymcmttmttmttmttmttmttmttmttmttmttmttmttmt 995
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: August 27, 2000 16:02:24
Job Name: 16150 seq



 WISEMAN (TM)

Release 3.1A John F. Collins, BioComputing Research Unit,
 Copyright (c) 1998 1999 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein : protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:53:48 2000; MaxRun time 10.75 seconds

Tabular output not generated. 574,123 Million cell updates/sec

Title: >US-09-257-585-4

Description: (1-89) from US09257585.pep

Sequence: 1 MKVQATFAITAIATFAATFAA.....DKSVLSDFISAVADHESC 89

Scoring table: PAM 150

Searched: 225878 seqs, 6934122 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database:

spiremb12
 1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
 5:sp-invertebrate 6:sp-mammal 7:sp-moll 8:sp-plant
 9:sp-phage 10:sp-plant 11:sp-protozoa 12:sp-unclassified
 13:sp-vertebrate 14:sp-virus

Statistics: Mean 38.662; Variance 66.101; scale 0.585

Prod. No. is the number of results predicted by change to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Prod. No.
1	93	14.7	723	5	045470	19669.12 PROTEIN.
2	92	14.5	633	14	032067	FE E31.
3	91	14.4	1275	4	015057	K1A0349 (FRAGMENT).
4	91	14.4	1363	5	018419	C3484.2 PROTEIN.
5	91	14.4	1661	5	006166	MATURE PARASITIC INFECT
6	90	14.2	1201	5	018978	NA-K-CT1 CONPARANOMETER
7	90	14.2	1510	5	025920	MATURE PARASITIC INFECT
8	89	14.0	353	2	092NMD	3-HYDROXY-3-METHYLOL
9	88	13.9	875	5	019045	FO3A3.6 PROTEIN.
10	87	13.7	219	14	072729	PGT1 PROTEIN.
11	87	13.7	301	1	057964	101AA LONG HYPOTHECA
12	87	13.7	326	3	0998G9	MATING TYPE PROTEIN MA
13	87	13.7	391	1	058489	301AA LONG HYPOTHECA
14	87	13.7	592	10	092857	PUTATIVE MULTISPANNING
15	87	13.7	729	2	051162	FE-PRIMATED PROTEIN S
16	87	13.7	922	2	094593	YMOA PROTEIN (MEMBER O
17	87	13.7	4017	14	098508	(PERDURE-115).
18	86	13.6	192	1	059618	192AA LONG HYPOTHECA
19	86	13.6	728	5	09XG34	Y37A1B.14 PROTEIN.
20	85	13.4	344	5	09XTB9	H22D14.1 PROTEIN.

21	85	13.4	494	5	07436	FL4A5.1 PROTEIN
22	85	13.4	492	10	04728	GLYCOPROTEIN
23	85	13.4	808	4	000602	COLLI/INIBIT
24	84	13.2	418	4	007995	PUTATIVE MIP
25	84	13.2	789	6	000967	COOH-TERP
26	83	13.1	167	5	02709	GLYCIN XIT
27	83	13.1	815	5	007834	PROTEIN W PE
28	83	13.1	726	5	008760	ORE YR103A
29	82	12.9	296	1	013076	ZAL1
30	82	12.9	512	5	018709	SIMILAR
31	82	12.9	508	5	019723	SMILAR
32	82	12.9	497	2	00554	GTAA D-N
33	81	12.8	154	1	000767	TI408-83
34	81	12.8	164	14	041204	GLYCOPROTEIN
35	81	12.8	184	13	061265	GLYCOPROTEIN
36	81	12.8	214	2	005538	HYPOHELI
37	81	12.8	260	14	005085	S-PROTEIN
38	81	12.8	306	2	002701	HYPOHELI
39	81	12.8	403	2	007565	HYPOHELI
40	81	12.8	449	14	007847	VIRUS E5
41	81	12.8	449	14	008510	VIRUS E5
42	81	12.8	449	14	008510	RNA. 05
43	80	12.4	179	14	009922	HYMILIC
44	80	12.4	426	5	025675	HYMILIC
45	80	12.4	718	14	008418	SV1

ALIGNMENTS

RESULT	ID	DATE	DESCRIPTION	PROT	LEN	AA
1	045470	01-JUN-1998	(TEMPORARY) Created			
2	01-JUN-1998	(TEMPORARY) Last sequence				
3	01-JUN-1998	(TEMPORARY) Last sequence				
4	01-JUN-1998	(TEMPORARY) Last sequence				
5	01-JUN-1998	(TEMPORARY) Last sequence				
6	01-JUN-1998	(TEMPORARY) Last sequence				
7	01-JUN-1998	(TEMPORARY) Last sequence				
8	01-JUN-1998	(TEMPORARY) Last sequence				
9	01-JUN-1998	(TEMPORARY) Last sequence				
10	01-JUN-1998	(TEMPORARY) Last sequence				
11	01-JUN-1998	(TEMPORARY) Last sequence				
12	01-JUN-1998	(TEMPORARY) Last sequence				
13	01-JUN-1998	(TEMPORARY) Last sequence				
14	01-JUN-1998	(TEMPORARY) Last sequence				
15	01-JUN-1998	(TEMPORARY) Last sequence				
16	01-JUN-1998	(TEMPORARY) Last sequence				
17	01-JUN-1998	(TEMPORARY) Last sequence				
18	01-JUN-1998	(TEMPORARY) Last sequence				
19	01-JUN-1998	(TEMPORARY) Last sequence				
20	01-JUN-1998	(TEMPORARY) Last sequence				


```

RESULT 14
ID 092PS7 PRELIMINARY: PRT: 592 AA.
AC 092PS7:
DT 01-MAY-1999 (TEMBLE: 10, created)
DI 01-MAY-1999 (TEMBLE: 10, last sequence update)
DI 01-MAY-1999 (TEMBLE: 10, last annotation update)
DE POTATIVE MULTISPANNING MEMBRANE PROTEIN.
GN F14H20.4.
OS Arabidopsis thaliana (Mouse-ear cress):
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
OC core eudicots: Rosidae: eurosids 1: Brassicales: Brassicaceae:
OC Arabidopsis.
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN:CV: COLUMBIA:
RA LIN X., KAUL S., SHIA I.P., FULLI C.Y., SHEN M., VANAKIS S.E.,
RA HARNSTEAD M.E., MASON T.M., ROMAN C.L., KONING C.M., BENITO M.,
RA CARREIRA A.J., GREASY J.H., HOBEL C.R., TOWN C.D., NIERMAN W.C.,
RA FRASER C.M., VENTER J.C.:
RT "Arabidopsis thaliana chromosome II BAC F14H20 genomic sequence."
RL Submitted (MAR 1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006532; A002090.1;
SD SEQUENCE 592 AA: 68048 MW: A0675F42 CKC32;

Query Match 13.7%: Score 87; Lb 10; Length 592;
Best Local Similarity 26.4%: Prol. No. 6,939-01;
Matches 15; Conservative 14; Mismatches 27; Indels 1; Gaps 1;

Db 1 MRUPTLFLGALPFGAGIVASDHVKKDQSVLYANKVQPHNPSITYRF 57
yy 1 MKVQATVATITAIYAAPPANANWYVITLPPKQTVNINITYAHNE-EITTEYKYF 56

RESULT 15
ID 051162 PRELIMINARY: PRT: 720 AA.
AC 051162:
DT 01-NOV-1996 (TEMBLE: 01, created)
DI 01-NOV-1996 (TEMBLE: 01, last sequence update)
DI 01-NOV-1996 (TEMBLE: 08, last annotation update)
DE PE-REGULATED PROTEIN B PRECURSOR.
GN PRPB.
OS Neisseria meningitidis.
OC Bacteria: Proteobacteria: beta subdivision: Neisseriaceae: Neisseria.
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN:444/76;
RA MEDLINE: 96009786;
RA PETERSSON A., MAAS A., VAN WASSENAAR D., VAN DER LUY P., TOMMASSEN J.;
RT "Molecular characterization of Prpb, the 70-kilodalton iron-regulated
RT outer membrane protein of Neisseria meningitidis."
RL Infect. Immun. 63:4181-4184(1995).
DR EMBL: X89755; CAA61902.1;
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 720
SD SEQUENCE 720 AA: 79611 MW: 4F79FA16 CRC32;

Query Match 13.7%: Score 87; Lb 2; Length 720;
Best Local Similarity 14.5%: Prol. No. 6,939-01;
Matches 20; Conservative 15; Mismatches 18; Indels 5; Gaps 4;

Db 186 KENNEDIESYNDEKDYFAKGFVNNGKTVYSA-LDFSYLA-KIGTFEEDD 241
yy 32 KDCIVNNIYIANNHETTFEYKVF--VNCA-CHYTPVILIPKSVLSQPTISAVARD 86

```

Search completed: Fri Aug 25 17:54:52 2000
Job time : 64 secs.



[illegible]

```

ENTRY      3
TITLE      A55015 #type complete
           bumetanide-sensitive Na(+)-K(+)-ATPase mouse
ORIGINISM
DATE
ACCESSIONS
REFERENCE
AUTHORS    Delipati, E.; Kanchanar, M.L.; Heler, D.K.; Hebert, S.C.;
           Gillars, S.P.
#journal   J. Biol. Chem. (1994) 269:56677-56683
#title     Molecular cloning and chromosomal localization of a putative
           basolateral Na(+)-K(+)-ATPase cotransporter from mouse inner ear
           mucillary collecting duct (mimcd-3) cells.
#citations  references MIMD:9501474
#accession A55015
#status    Preliminary
#molecule_type mRNA
#residues  1-1205 ##label DEL
#cross-references GSI:013174; NID:q560007; PID:q560008
SUMMARY    #length 1205 #molecular_weight 130kDa #cvsespm 6188
query Match          14.7% Score 93; DB 2; Length 1205;
best Local Similarity 25.5%; Pred. No. 4,13e+01;
Matches 13; Conservative 13; Mismatches 23; Indels 2; Gaps 2
db 1028 LILLIPUT11LRKMNDPCNRYVLCQCNINRDIHDPKAMALISKEPIDESD 1078
      I + + : : : 111 1 : : : 111 : : : : :
      QY 10 LIALAVFPAAANA-WKDCIIGIKYKHIVNNITYTANNEETIEFYKV-FVN 58
RESULT      4
ENTRY       4
TITLE       A57187 #type complete
           bumetanide-sensitive Na-K-Cl cotransporter - human
ORGANISM    HUMAN
DATE        16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change
           28-Feb-1997
ACCESSIONS  A57187
REFERENCE    Payne, J.A.; Xu, J.C.; Haas, M.; Lytle, C.Y.; Ward, D.;
#authors     Foltusht, B.
#journal     J. Biol. Chem. (1996) 270:17977-17985
#title       Primary structure, functional expression, and chromosomal
           localization of the bumetanide-sensitive Na-K-Cl
           cotransport in human colon.
#cross-references MIMD:9545597
#accession  A57187
#status     preliminary

```

```
##motif_type MRNA
##residues 1-1212 ##label RES
#Accession EMBL:U0946, F01470001, U09470002
GENEFON
#GDB NR001
#Ref_Fc1118 5123-4
#length 1212 #molecular_weight 13144 #molecular_2210

Query Match      14.7% Score 94; DB Z; Length 1212;
Post. Local Similarity 25.5%; Pred. No. 2.1e-01;
Matches 13; Conservative 13; Mismatches 23; Indels 2; Gaps 2;

126 1045 LLLIYLLTEKKKMKLEVEFQKLNINIEPAMATLSKEPIESD 1085
       | : : : | : : : | : : : | : : : | : : : |
02   10 IIAAVFDVANA-WRDTIGGYKQGVNNVIANYAREHILIRKV-VVN 58
```

[illegible]

Sun Aug 27 16:02:02 2000

US-09-257-585-4.rpr

Page 6

2007 Match	17443	Score 65	Dr 2	Length 260
Best local similarity	39.4%	Hyd. No. 271600		
Matches	13	Conservative	7	Mismatches 11
				Indels

2, Gaps 2,

159 IQUALIT-VALGSYTALQAW-EYIDAFETIAD 189

35 VÔAT'ATILAI AAYEPANAWKIDC I ÖRYKIDC) 35 QY

Search completed: Fri Aug 25 17:52:46 2000
Job time : 19 secs.


```

or send an email to license@sb.slb.ch).
CC -----
DB EMBL: U01474; AAC57784.1.1.
DB MIM: MG1:101924; SL012A2.
DB PIR: P00324; aa_permease.1.
DB PRINTS: PR01207; NAKLTNSPRT.
DB PRINTS: PR01208; NAKLTNSPRT.
KW Transport: Transmembrane; Glycoprotein.
CC -----
FT DOMAIN 1 278 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 279 299 POTENTIAL.
FT DOMAIN 300 302 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 303 323 POTENTIAL.
FT DOMAIN 324 360 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 361 481 POTENTIAL.
FT DOMAIN 382 403 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 404 424 POTENTIAL.
FT DOMAIN 425 428 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 429 449 POTENTIAL.
FT DOMAIN 450 479 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 480 500 POTENTIAL.
FT DOMAIN 501 513 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 514 534 POTENTIAL.
FT DOMAIN 535 564 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 585 605 POTENTIAL.
FT DOMAIN 606 650 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 651 671 POTENTIAL.
FT DOMAIN 672 692 POTENTIAL.
FT TRANSSEM 693 709 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 710 740 POTENTIAL.
FT TRANSSEM 731 892 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 913 913 POTENTIAL.
FT TRANSSEM 914 1205 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 940 107 POTENTIAL.
FT TRANSSEM 546 546 POTENTIAL.
FT CARBOHYD 555 575 POTENTIAL.
SU SEQUENCE 1208 AA; 130649 MW; A93E29A468377125 CR654;

Query Match 14.78; Score 94; DB 1; Length 1205;
Host Local Similarity 25.58; Prod. No. 3.970-02;
Matches 13; Conservative 13; Mismatches 23; Indels 2; Gaps 2;

DB 1028 LTLIPYLTKRKWKPKFVETGKINSIDHPRAMATLSKPIEDSD 1078
| : : : : | : : : : | : : : : | : : : : |
QY 10 LIALAAVFPANA-WKDCITQRYKGDVNNIVYANNEETIEYKV-FVN 58

RESULT 4
ID NKCL_HUMAN STANDARD: PRT: 1212 AA.
AC P55011;
DI 01-OCT-1996 (Rel. 34, created)
DI 01-OCT-1996 (Rel. 34, last sequence update)
DI 01-NOV-1997 (Rel. 35, last annotation update)
DE HEMETANIDE-SENSITIVE SODIUM-PROFASINUM-CHLORIDE COTRANSPORTER 1
DE (UNISOLATED NA-K CL SYM-PTER).
GN SLC12A2 (OR NKCL1).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE-COLON.
RX MEDLINE: 95355397.
RA Payne J.A., Xu J.-C., Haas M., Lytle Y.C., Ward D., Forbush B. 111;
RA "Primary structure, functional expression, and chromosomal
RA localization of the human iso-ionic-sensitive Na-K-Cl cotransporter in
RA human colon."
RT J Biol Chem 270:17977-17985(1995).
CC -1- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
CC MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION. PLAYS A VITAL ROLE
CC IN THE REGULATION OF TONGUE PALATE AND GILL VOLUME.
CC -1- SUBCELLULAR LOCATION: INTRACELLULAR MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES.
CC -1- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.

```

```

CC -----
CC This SWISS-PROT entry is copyrighted. It is produced
CC between the Swiss Institute of Bioinformatics and
CC the European Bioinformatics Institute. There are
CC use by non-profit institutions as long as it is
CC modified and this statement is not removed. User
CC entities requires a license agreement (See file //a
CC or send an email to license@sb.slb.ch).
CC -----
DB EMBL: U03246; AAC50561.1.
DB MIM: 600840.1.
DB PIR: P00324; aa_permease.1.
DB PRINTS: PR01207; NAKLTNSPRT.
DB PRINTS: PR01208; NAKLTNSPRT.
KW Transport: Transmembrane; Glycoprotein.
CC -----
FT DOMAIN 1 92 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 93 113 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 114 285 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 286 406 POTENTIAL.
FT DOMAIN 407 409 POTENTIAL.
FT TRANSSEM 410 430 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 431 466 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 467 487 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 488 411 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 411 441 POTENTIAL.
FT DOMAIN 442 442 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 443 456 POTENTIAL.
FT DOMAIN 457 496 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 487 507 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 508 520 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 521 541 POTENTIAL.
FT DOMAIN 542 561 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 562 612 POTENTIAL.
FT DOMAIN 613 652 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 653 672 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 673 704 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 705 724 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 725 745 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 746 1212 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 93 111 POTENTIAL.
FT TRANSSEM 188 194 POTENTIAL.
FT CARBOHYD 168 194 POTENTIAL.
SU SEQUENCE 1212 AA; 131447 MW; AB0F174AC7A2AE7;

Query Match 11.74; Score 93; DB 1; Length 1205;
Host Local Similarity 25.58; Prod. No. 3.970-02;
Matches 13; Conservative 13; Mismatches 23;

DB 1045 LTLIPYLTKRKWKPKFVETGKINSIDHPRAMATLSKPIEDSD 1078
| : : : : | : : : : | : : : : | : : : : |
QY 10 LIALAAVFPANA-WKDCITQRYKGDVNNIVYANNEETIEYKV-FVN 58

RESULT 4
ID NIFL_FOXPV STANDARD: PRT: 647 AA.
AC Q72907;
DI 15-FEB-2000 (Rel. 39, created)
DI 15-FEB-2000 (Rel. 39, last sequence update)
DI 15-FEB-2000 (Rel. 39, last annotation update)
DE NUCLEOSIDE TRIPHOSPHATASE 1 (NTP1).
DE PHOSPHOHYDROLASE 1 (NTP1).
GN Pp-D11.
OS Fowlpox virus.
OC Viruses; dsDNA viruses; ssRNA status: positive;
OC Arthropods.
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN FPV / BP-44.
RA Skinner M.A.;
RA Submitted (MAY 1999) to the EMBL Data Bank.
CC -1- FUNCTION: SERVES TWO PATHS IN TRANSPORTING
CC WITH VITAL TERMINAL N-GLYCOSYLATION.
CC OF DURING CONTINUOUS NUCLEOTIDE RNA FROM THE P.

```

```

01 IT ACTS BY ITSELF AS A POLYMERASE ELONGATION FACTOR TO FACILITATE
02 REACHING OF INTERFASAL STAGES (BY SIMILARITY).
03 -1- CATALYTIC ACTIVITY: NTP + H2O -> NMP + GTP/GDP/GTP-P.
04 -1- SIMILARITY: BELONGS TO THE NTP 1 SUBFAMILY OF HELICASES.
05
06 THIS SWISS-PROT entry is copyrighted. It is reproduced through a collaboration
07 between the Swiss Institute of Bioinformatics and the EMBL outstation at
08 the European Bioinformatics Institute. There are no restrictions on its
09 use by non-profit institutions as long as its content is in no way
10 modified and this statement is not removed. Usage by and for commercial
11 entities requires a license agreement (see http://www.ebi.ac.uk/infocentre/copyright.htm
12 or send an email to license@ebi.ac.uk).
13
14 -----
15 DR EMBL: AJ005163; CAA06401.1;
16 DR PDB: P600176; SNF2_N; 1;
17 DR PDB: P600271; helicase_C; 1;
18 KW Helicase; ATP binding; transcription; late protein.
19 FT DOMAIN 397 483 SNF2_N
20 FT NP_BIND 56 64 ATP (BY SIMILARITY).
21 FT SITE 142 145 DECK BOX.
22 SQ SEQUENCE 637 AA: 73774 MW: 983482.05147998 CYS64.
23
24 Query Match 14.5% Score 92; DB 1; Length 647;
25 Best Local Similarity 41.0% Pred. No. 5,74e-02;
26 Matches 13; Conservative 11; Mismatches 15; Indels 3; Gaps 3;
27
28 DB 108 KCTD11MNYVHQDHPHAKPEFINKS-NVKS-KLPLDLPDQHP-147
29 UY 24 KDCIIGYKRGIVNRIYIANREHETIEEVKVFVN-ENCIHFY 64
30
31
32 RESULT 6
33 ID CG23-YEAST STANDARD; PRT; 427 AA.
34 ID P24979;
35 DT 01-MAR-1992 (Ref. 21, created)
36 DT 01-OCT-1993 (Ref. 22, last sequence update)
37 DT 15-JUL-1999 (Ref. 38, last annotation update)
38 DT G2M10000 SPECIALTY CYCLIN 4
39 CN CUB3 OR YDL155W OR D1539
40 OS Saccharomyces cerevisiae (Baker's yeast)
41 GN Eukaryotic Cyclin A/Ascomycota1; Saccharomycetes; Saccharomycelales;
42 GN Saccharomycetaceae; Saccharomycetes.
43 GN [1]
44 SEQUENCE FROM N.A.
45 STRAIN S288c;
46  MDLINE: 93051123.
47  Pritchardson II., Lew P.J., Benze M., Sugimoto K., Reed S.I.;
48  "Cyclin B homologs in Saccharomyces cerevisiae function in S phase
49  and in G2."
50  Genes Dev. 6:2021-2034(1992).
51 [2]
52 SEQUENCE FROM N.A.
53 MDLINE: 92485891.
54  Fitch I., Falmagne C., Strana W., Amor A., Nasmyth K., Gietzsch L.;
55  Byers B., Fletcher B.;
56  "Characterization of four B-type cyclin genes of the budding yeast
57  Saccharomyces cerevisiae."
58  Mol. Biol. Cell 3:805-818(1992).
59 [3]
60 SEQUENCE FROM N.A.
61 MDLINE: 91191954.
62  Strana W., Kobitsch H., Priva C., Schuster T., Fitch I., Fletcher A.B.,
63  Nasmyth K.;
64  "The role of CUB3 and cyclins during mitosis in the budding yeast S.
65  cerevisiae."

```

2L COL1 65:45-161(1991).
 3L -1- TERMINAL: ESSENTIAL FOR THE CONTROL OF THE CELL CYCLE AT THE G2/M
 3M (MAYNARD) TRANSITION. INTERACTS WITH THE G2/M PROMOTER KINASE TO
 3N FORM MPF. G2/M CYCLIN ACCUMULATE SPECIFICALLY DURING G2 AND ARE
 3O ACCORDINGLY DESTROYED AT MITOSIS.
 3P
 3Q -1- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
 3R
 3S THIS SWISS PROTEIN is copyright. It is produced through a collaboration
 3T between the Swiss Institute of Bioinformatics and the EMBL collaboration.
 3U the European Bioinformatics Institute. There are no restrictions on its
 3V use. If you prefer that it not be used as a contact us in any way
 3W modified and this statement is not removed. Usage by and for commercial
 3X entities requires a license agreement (see <http://www.isb.scrib.ch/author/cy>
 3Y or send an email to license@isb.scrib.ch).
 3Z
 4L EMBL: X69425; CAA49201.1;
 4M EMBL: M80302; AAA44765.1;
 4N EMBL: X97751; CAA66336.1;
 4O EMBL: Z74203; CAA98729.1;
 4P EMBL: S14167; S14167.
 4Q EMBL: A46228; A46228.
 4R EMBL: A60548; A60548.
 4S HSPD: P30274; 1V1N.
 4T SDD: L0000351; CLB3.
 4U PFAM: PF00134; CYCLIN. 1.
 4V PROSITE: PS00292; CYCLINS; 1.
 4W CYCLIN1: Cell cycle; Cell division; Mitosis; Multigene family.
 4X COMBILIT 48 48 R 3 D (1N REP. 2 AND CAA98729).
 4Y SEQUENCE 427 AA: 49516 MW: 205484787676728 CRR64.
 5L
 5M
 5N
 5O
 5P
 5Q
 5R
 5S
 5T
 5U
 5V
 5W
 5X
 5Y
 5Z
 6L
 6M
 6N
 6O
 6P
 6Q
 6R
 6S
 6T
 6U
 6V
 6W
 6X
 6Y
 6Z
 7L
 7M
 7N
 7O
 7P
 7Q
 7R
 7S
 7T
 7U
 7V
 7W
 7X
 7Y
 7Z
 8L
 8M
 8N
 8O
 8P
 8Q
 8R
 8S
 8T
 8U
 8V
 8W
 8X
 8Y
 8Z
 9L
 9M
 9N
 9O
 9P
 9Q
 9R
 9S
 9T
 9U
 9V
 9W
 9X
 9Y
 9Z
 10L
 10M
 10N
 10O
 10P
 10Q
 10R
 10S
 10T
 10U
 10V
 10W
 10X
 10Y
 10Z
 11L
 11M
 11N
 11O
 11P
 11Q
 11R
 11S
 11T
 11U
 11V
 11W
 11X
 11Y
 11Z
 12L
 12M
 12N
 12O
 12P
 12Q
 12R
 12S
 12T
 12U
 12V
 12W
 12X
 12Y
 12Z
 13L
 13M
 13N
 13O
 13P
 13Q
 13R
 13S
 13T
 13U
 13V
 13W
 13X
 13Y
 13Z
 14L
 14M
 14N
 14O
 14P
 14Q
 14R
 14S
 14T
 14U
 14V
 14W
 14X
 14Y
 14Z
 15L
 15M
 15N
 15O
 15P
 15Q
 15R
 15S
 15T
 15U
 15V
 15W
 15X
 15Y
 15Z
 16L
 16M
 16N
 16O
 16P
 16Q
 16R
 16S
 16T
 16U
 16V
 16W
 16X
 16Y
 16Z
 17L
 17M
 17N
 17O
 17P
 17Q
 17R
 17S
 17T
 17U
 17V
 17W
 17X
 17Y
 17Z
 18L
 18M
 18N
 18O
 18P
 18Q
 18R
 18S
 18T
 18U
 18V
 18W
 18X
 18Y
 18Z
 19L
 19M
 19N
 19O
 19P
 19Q
 19R
 19S
 19T
 19U
 19V
 19W
 19X
 19Y
 19Z
 20L
 20M
 20N
 20O
 20P
 20Q
 20R
 20S
 20T
 20U
 20V
 20W
 20X
 20Y
 20Z
 21L
 21M
 21N
 21O
 21P
 21Q
 21R
 21S
 21T
 21U
 21V
 21W
 21X
 21Y
 21Z
 22L
 22M
 22N
 22O
 22P
 22Q
 22R
 22S
 22T
 22U
 22V
 22W
 22X
 22Y
 22Z
 23L
 23M
 23N
 23O
 23P
 23Q
 23R
 23S
 23T
 23U
 23V
 23W
 23X
 23Y
 23Z
 24L
 24M
 24N
 24O
 24P
 24Q
 24R
 24S
 24T
 24U
 24V
 24W
 24X
 24Y
 24Z
 25L
 25M
 25N
 25O
 25P
 25Q
 25R
 25S
 25T
 25U
 25V
 25W
 25X
 25Y
 25Z
 26L
 26M
 26N
 26O
 26P
 26Q
 26R
 26S
 26T
 26U
 26V
 26W
 26X
 26Y
 26Z
 27L
 27M
 27N
 27O
 27P
 27Q
 27R
 27S
 27T
 27U
 27V
 27W
 27X
 27Y
 27Z
 28L
 28M
 28N
 28O
 28P
 28Q
 28R
 28S
 28T
 28U
 28V
 28W
 28X
 28Y
 28Z
 29L
 29M
 29N
 29O
 29P
 29Q
 29R
 29S
 29T
 29U
 29V
 29W
 29X
 29Y
 29Z
 30L
 30M
 30N
 30O
 30P
 30Q
 30R
 30S
 30T
 30U
 30V
 30W
 30X
 30Y
 30Z
 31L
 31M
 31N
 31O
 31P
 31Q
 31R
 31S
 31T
 31U
 31V
 31W
 31X
 31Y
 31Z
 32L
 32M
 32N
 32O
 32P
 32Q
 32R
 32S
 32T
 32U
 32V
 32W
 32X
 32Y
 32Z
 33L
 33M
 33N
 33O
 33P

```

KA  MEDLINE: 94073773.
KB  Smith M.C., Murray B.L.
KT  *Sequence analysis of the beta-lactamase repressor from
KI  Staphylococcus aureus and hybridization studies with two
KL  beta-lactamase-producing isolates of Enterococcus faecalis."
KM  Antimicrob. Agents Chemother. 36:2265-2269(1992).
KN  -1- FUNCTION: REPRESSION OF BETA-LACTAMASE.
KO  -1- SIMILARITY: BELONGS TO THE BLA1/MBL FAMILY OF TRANSCRIPTIONAL
K1  REGULATORS.
K2
K3
K4  THIS SWISS-PROT entry is copyright. It is produced through a collaboration
K5  between the Swiss Institute of Bioinformatics and the EMBL outstation
K6  at the European Bioinformatics Institute. There are no restrictions on its
K7  use by non-profit institutions as long as its content is in no way
K8  modified and this statement is not removed. Usage by and for commercial
K9  entities requires a license agreement (see http://www.ebi.ac.uk/infocentre/copyright.htm
LA  or send an email to license@isb.stg.ch).
LB
LC  EMBL: X5734; CAA36951.1.
LD  EMBL: M62650; AAA26603.1.
LE  EMBL: M62476; AAA26605.1.
LF  PIR: S34446; S34446.
LG  PIR: S11782; S11782.
LH  Transposable element; Transcription regulation; Repressor;
LI  DNA binding; Plasmid.
LJ  SOURCE: 126 AA; 14875 MW; E075948A5183117 CRC64;
LK
LM  Query Match: 13.9%; Score 88; DB 1; Length 126;
LN  Best Local Similarity 33.3%; Pred. No. 2 426-01;
LO  Matches 12; Conservative 12; Mismatches 9; Indels 4; Gaps 3,
LP
LQ  55 Kc-111KRSYSLNY-PTSSNADNDQAMAMIANPIN 89
LR  1-11111111111111111111111111111111
LS  24 KMLTGYEYEEYEVNITVAN EREETTYEEVEYH 58
LT
LU  RESULT 8
LV  ID NFN.SYND8 STANDARD PRI: 461 AA.
LW  AC 007456;
LX  DT 15-JUL-1999 (Rel. 48, Created)
LY  DT 15-JUL-1999 (Rel. 38, Last sequence update)
LZ  DT 15-JUL-1999 (Rel. 48, Last annotation update)
MA  NITROGENASE IRON-MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN NFN.
MB  NFN.
MC  Synthesized as specified by E.C. 7.1.1 (Molybdenum FET 8801).
MD  Bacteroides (gamma) Bacteroides thuringiensis B. thuringiensis
ME  [1]
MF  SEQUENCE FROM N.A.
MG  Chen H.M.;
MH  "NITENXM OPERON."
MI  Submitted (MAY-1997) to the EMBL/Genbank/NCBI databases
MJ  -1- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF
MK  THE PROSTHETIC GROUP OF NITROGENASE (FETMO COFACTOR).
ML  -1- PATHWAY: FETMO COFACTOR BIOSYNTHESIS.
MN  -1- SIMILARITY: BELONGS TO THE NIF-H/NIFV/NIFU/NIFN FAMILY.
MO
MP  This SWISS-PROT entry is copyright. It is produced through a collaboration
MQ  between the Swiss Institute of Bioinformatics and the EMBL outstation
MR  at the European Bioinformatics Institute. There are no restrictions on its
MS  use by non-profit institutions as long as its content is in no way
MT  modified and this statement is not removed. Usage by and for commercial
MU  entities requires a license agreement (see http://www.ebi.ac.uk/infocentre/copyright.htm
MV  or send an email to license@isb.stg.ch).
MW
MX  EMBL: AF003700; AAC35195.1.
MY  PIR: P001148; oxidized_nitro_1.
MZ  PROSITE: PS00699; NITROGENASE_L1_1.
NA  PROSITE: PS00940; NITROGENASE_L2_2; PALSE_NMC.
NB  NITROGEN FIXATION.
NC  SFEQWVF 461 AA; 50377 MW; 04776CAVAVBPQAFV CEC64;
ND
NE  Query Match: 13.6%; Score 86; DB 1; Length 461;
NF  Best Local Similarity 33.3%; Pred. No. 4 896-01;

```

[illegible]

Search completed: Fri Aug 25 17:53:31 2000
 Job time : 27 secs.

```

RESULT 15
ID CARB_YEAST STANDARD: PRT: 1118 AA.
AC P03965:
OT 23-OCT-1986 (rel. 02, created)
OT 01-OCT-1986 (rel. 02, last sequence update)
OT 01-OCT-1996 (rel. 34, last annotation update)
DE CARBAMOYL-PHOSPHATE SYNTHASE, ARGININE-SPECIFIC, LARGE CHAIN
DE (EC 6.3.5.5) (ARGININE-SPECIFIC CARBAMOYL-PHOSPHATE SYNTHETASE,
DE AMMONIA CHAIN).
ON CPA2 OR YJRI09C OR J2002.
OS Saccharomyces cerevisiae (baker's yeast).
OT Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OT Saccharomycetaceae; Saccharomyces.
AN 111
AN SEQUENCE FROM N.A.
AX MEDLINE: 84061899.
KA Lusty C.J., Widgren E.E., Broglie K.E., Nymoya H.:
KT Yeast carbamyl phosphate synthetase: structure of the yeast gene and
KT homology to Escherichia coli carbamyl phosphate synthetase."
KL J. Biol. Chem. 258:14466-14477(1983).
KN 12]
AP SEQUENCE FROM N.A.
KA Ramezani Rad M., Kirchbach L., Hollenberg C.P.:
KL Submitted (SEP 1995) to the EMBL/GenBank/DBJ databases.
KN 13]
AP SEQUENCE OF 1-859 FROM N.A.
KA Rose M., Koetter P., Entian K.D.:
KL Submitted (SEP 1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 2 ATP + GLUTAMINE + CO(2) -> H(2)O + 2 ADP +
CC ORTHOPHOSPHATE + GLUTAMATE + CARBAMOYL PHOSPHATE.
CC -1- PATHWAY: BIOSYNTHESIS OF ARGININE.
CC -1- SUBUNIT: COMPOSED OF TWO CHAINS: THE SMALL (OR GLUTAMINE) CHAIN
CC PROMOTES THE HYDROLYSIS OF GLUTAMINE TO AMMONIA, WHICH IS USED
CC BY THE LARGE (OR AMMONIA) CHAIN TO SYNTHESIZE CARBAMOYL PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- MISCELLANEOUS: IN EUKARYOTES THIS ENZYME IS SYNTHESIZED BY TWO
CC PATHWAY-SPECIFIC (ARGININE AND PYRIMIDINE) UNDER SEPARATE CONTROL.
CC -1- SIMILARITY: HOMOLOGOUS TO THE PROKARYOTE ENZYME.
CC
CC This SWISS-Prot entry is copyright. It is produced through a collation
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/submit/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: K01178; AAA66902.1;
CC EMBL: Z49609; CAA89639.1;
CC PIR: A01199; SYRSCP.
CC HSSP: P00968; 1J08.
CC SCOP: 10000400; CPA2.
CC PFAM: PF00289; cPase_L_cblan_2.
CC PRINTS: PR00098; cPASE.
CC PROSITE: PS00866; cPASE_1; 2.
CC PROSITE: PS00867; cPASE_2; 2.
CC Arginine biosynthesis; lidase; duplication; ATP binding.
CC
CC REPEAT 551 1107
CC NP_BIND 174 229 ATP (POTENTIAL).
CC NP_BIND 421 471 ATP (POTENTIAL).
CC
CC SEQUENCE 1118 AA; 123914 MW; 887FAAF00A007674 CAC66;
CC
CC
CC QUOTE Match 12.98; Score 82; DR 1; Length 1118;
CC Best Local Similarity 28.48; Prod. No. 1.93e+00;
CC Matches 15; Conservative 16; Mismatches 20; Indels 2; Gaps 2;
CC
CC 702 DSLVDVDPPEWSELSVEEAKIFASKV-N-YPLLIRPSVLSGAAKSVVNNPTEE 752
CC 45 DVNNIVTARNRPETITPEYKVFVNACHDVYVILLDRSVLSGDETSAYADNDE 87
  
```

MORSE

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Match_pp protein - protein database search, using Smith-Waterman algorithm
Program: 171 Aug 25 17:55:10 2000 Master time 3.23 seconds
422,475 Million cell updates/sec
Tabular output not generated

Title: >US-09-257-585-4
Description: (1-89) from US09257585.pcp
Perfect Score: 634
Sequence: 1 MKVQATFALLALAAFPAA.....DRVLSQDTSAVDDESC 89

Scoring table:
PAM 150
Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 04
Listing first 45 summaries

Database: a-issued
1:5A_00MH 2:5H_00MH 3:6_00MH 4:PGI_00MH 5:backfiles1

Statistics: Mean 26.514; Variance 111.214; Scale 0.238

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	DB	ID	Description	Prod. No.
No.	Score						
1	78	12.3	286	1	US-08-118-	Sequence 65, Applicant 4, 73e+01	
2	78	12.3	286	4	PGI 0593-0	Sequence 65, Applicant 4, 73e+01	
3	77	12.1	412	1	US-87-712	Sequence 3, Applicant 5, 63e+01	
4	76	12.0	203	2	US-08-846-	Sequence 1, Applicant 6, 70e+01	
5	76	12.0	231	2	US-08-902-	Sequence 6, Applicant 6, 70e+01	
6	76	12.0	296	1	US-09-116-	Sequence 6, Applicant 6, 70e+01	
7	76	12.0	296	1	US-08-507	Sequence 6, Applicant 6, 70e+01	
8	76	12.0	529	2	US-08-846-	Sequence 3, Applicant 6, 70e+01	
9	75	11.8	227	1	US-08-458-	Sequence 4, Applicant 7, 96e+01	
10	75	11.8	454	2	US-08-933-	Sequence 2, Applicant 7, 96e+01	
11	75	11.8	454	3	US-08-960-	Sequence 2, Applicant 7, 96e+01	
12	75	11.8	1245	1	US-08-304	Sequence 8, Applicant 7, 96e+01	
13	75	11.8	1245	2	US-08-611-	Sequence 8, Applicant 7, 96e+01	
14	75	11.8	1245	3	US-08-611-	Sequence 8, Applicant 7, 96e+01	
15	75	11.8	1245	4	US-08-611-	Sequence 8, Applicant 7, 96e+01	
16	74	11.7	228	1	US-09-143-	Sequence 3, Applicant 9, 45e+01	
17	74	11.7	252	1	US-08-392-	Sequence 3, Applicant 9, 45e+01	
18	74	11.7	262	1	US-09-330	Sequence 3, Applicant 9, 45e+01	
19	74	11.7	262	5	US-09-330	Sequence 3, Applicant 9, 45e+01	
20	73	11.5	224	1	US-07-921-	Sequence 2, Applicant 1, 12e+02	
21	73	11.5	224	1	US-08-165	Sequence 2, Applicant 1, 12e+02	
22	73	11.5	224	1	US-09-411	Sequence 2, Applicant 1, 12e+02	
23	73	11.5	224	1	US-07-745	Sequence 2, Applicant 1, 12e+02	

24	73	11.5	224	2	US-09-412-	Sequence 4,
25	73	11.5	224	2	US-08-971-	Sequence 4,
26	73	11.5	224	2	US-08-290-	Sequence 8,
27	73	11.5	224	4	PGI 0594-0	Sequence 2,
28	73	11.5	407	1	US-08-745-	Sequence 14
29	73	11.5	407	1	US-08-597-	Sequence 14
30	73	11.5	420	1	US-08-461-	Sequence 10
31	73	11.5	420	1	US-07-945-	Sequence 10
32	73	11.5	420	1	US-08-451	Sequence 10
33	73	11.5	420	1	US-08-452	Sequence 10
34	73	11.5	420	4	PGI 0593-0	Sequence 11
35	73	11.5	546	1	US-08-426-	Sequence 2,
36	73	11.5	546	1	US-08-454-	Sequence 2,
37	73	11.5	546	1	US-08-401-	Sequence 2,
38	73	11.5	546	1	US-07-999-	Sequence 2,
39	73	11.5	546	4	PGI 0593-0	Sequence 2,
40	73	11.5	2161	2	US-08-223-	Sequence 4,
41	73	11.5	2161	1	US-08-455-	Sequence 4,
42	73	11.5	2161	1	US-08-455-	Sequence 5,
43	73	11.5	2161	1	US-08-745	Sequence 2,
44	73	11.5	2161	2	US-08-411	Sequence 2,
45	73	11.5	2161	2	US-08-223-	Sequence 6,

ALIGNMENTS

RESULT 1
ID US-08-118-270-65 STAMARD: PRT: 284 MA
AC xxxxxx
DE Sequence 65, Applicant 05/08:18270
CC Patent No. 5508364
CC GENERAL INFORMATION:
CC APPLICANT: MORTON, Randall B.
CC TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS OF MATTER
CC NUMBER OF SEQUENCES: 48
CC CORRESPONDING ADDRESS:
CC ADDRESSEE: JEFFREY AND NIMMER
CC STREET: 419 SWEET STREET, N.W., SUITE
CC CITY: WASHINGTON
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20004
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC OPERATING SYSTEM: PC COMPATIBLE
CC SOFTWARE: PATENT IN PROCESS #1-0, Version
CC ATTORNEY: R. B. BELL, 207 5118-270
CC FILING DATE: 09-SEP-1994
CC PRIOR APPLICATIONS:
CC APPLICATION: 08-SEP-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: JOWNSON, ROBERT G.
CC REGISTRATION NUMBER: 34,043
CC REFERENCE: 207 5118-270, MORTON-2A
CC TELEPHONE: 202 747 5197
CC TELEFAX: 202 747 5628
CC TELEX: 248634
CC INFORMATION R.K. AND IN NO: 65:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 286 amino acids
CC TYPE: amino acid

CC STANDARDNESS: single
 CC MOLECULE TYPE: linear
 CC MOLECULE TYPE: peptide
 SO SEQUENCE 286 AA; 31983 MW; 434446 CN.

Query Match 12.48; Score 78; DB 1; Length 286;
 Best Local Similarity 43.48; Pred. No. 4.73e+01;
 Matches 13; Conservative 11; Mismatches 10; Indels 5; Gaps 4;

DB 89 LAVNAYDHYVA-ICHHPIHVPVIVSSRTZLV-GSMWAGREG 125
 48 IIEETKVFVNEAHP-VIVVILIDR-SVLSISPTSAVA 83

RESULT 2
 ID PCT-US93-08528-65 STANDARD; PRI: 286 AA.

AC xxxxxx

DE Sequence 65, Application PCT/US93/08528

CC General Information: PCT/US93/08528

CC APPLICANT: New York University

CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN

CC NUMBER OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF

CC CORRESPONDENCE ADDRESS:

CC ADDRESS: BROOKLYN AND NEWARK

CC STREET: 419 Seventh Street, N.W., Suite 300

CC CITY: Washington

CC STATE: D.C.

CC COUNTRY: USA

CC ZIP: 20004

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: PCT/US93/08528

CC FILING DATE: 09-SEP-1994

CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: US 07/443,736

CC FILING DATE: 10-SEP-1992

CC APPLICANT/AGENT INFORMATION:

CC NAME: Townsend, Kevin G.

CC REPRESENTATION NUMBER: 44,033

CC REFERENCE/DOCKET NUMBER: MURPHY-2 PCT

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 202-628-5197

CC TELEFAX: 202-727-3528

CC FAX: 248633

CC INFORMATION FOR SEQ ID NO: 65:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 286 amino acids

CC TYPE: amino acid

CC STRANDNESS: single

CC MOLECULE TYPE: linear

CC SEQUENCE 286 AA; 31983 MW; 434446 CN.

Query Match 12.48; Score 78; DB 4; Length 286;
 Best Local Similarity 43.48; Pred. No. 4.73e+01;
 Matches 13; Conservative 11; Mismatches 10; Indels 5; Gaps 4;

DB 89 LAVNAYDHYVA-ICHHPIHVPVIVSSRTZLV-GSMWAGREG 125
 48 IIEETKVFVNEAHP-VIVVILIDR-SVLSISPTSAVA 83

RESULT 3
 ID US-07-712-833A-4 STANDARD; PRI: 412 AA.
 AC xxxxxx

Query Match 12.18; Score 77; DB 1; Length 412;
 Best Local Similarity 29.58; Pred. No. 5.63e+01;
 Matches 13; Conservative 12; Mismatches 17; Indels 2; Gaps 2;

DB 47 DQSKPFRKSHSKDIAVSTFFMANTFATMVGSPKQGVSV 90
 25 DCLIGRYKDS-DVNNITYTANRNEITIEEYVNE-AHPYPV 66

Sequence 3, Application US/07712833A
 Patent No. 5175101

AC xxxxxx

DE Sequence 3, Application US/07712833A

CC General Information: US/07712833A

CC APPLICANT: GOTT, Friedrich

CC APPLICANT: SEEBER, Stefan

CC TITLE OF INVENTION: RECOMBINANT RESTRICTION ENZYME SAU3AI

CC NUMBER OF SEQUENCES: 5

CC CORRESPONDENCE ADDRESS:

CC ADDRESS: Armstrong, Nikaido, Marumoto, Kikuchi &

CC STREET: 1725 K Street N.W., Suite 1000

CC CITY: Washington D.C.

CC COUNTRY: United States of America

CC ZIP: 20006

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/07712,833A

CC FILING DATE: 19910610

CC CLASSIFICATION: 530

CC PRIORITY APPLICATION DATA:

CC APPLICATION NUMBER: DE P 4018441.2

CC FILING DATE: 08-JUN-1990

CC APPLICANT/AGENT INFORMATION:

CC NAME: Murray, Robert B.

CC REPRESENTATION NUMBER: 22,980

CC REFERENCE/DOCKET NUMBER: 910808

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (202) 659-2940

CC TELEFAX: (202) 887-0357

CC FAX: 440142

CC INFORMATION FOR SEQ ID NO: 3:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 412 amino acids

CC TYPE: AMINO ACID

CC MOLECULE TYPE: linear

CC SEQUENCE 412 AA; 47294 MW; 868600 CN;

Query Match 12.18; Score 77; DB 1; Length 412;
 Best Local Similarity 29.58; Pred. No. 5.63e+01;
 Matches 13; Conservative 12; Mismatches 17; Indels 2; Gaps 2;

DB 47 DQSKPFRKSHSKDIAVSTFFMANTFATMVGSPKQGVSV 90
 25 DCLIGRYKDS-DVNNITYTANRNEITIEEYVNE-AHPYPV 66

Sequence 4, Application US/08836442
 Patent No. 5990293

AC xxxxxx

DE Sequence 4, Application US/08836442

CC General Information: US/08836442

CC APPLICANT: DOHERTY, Andrew, J.P.

01
XX
DE
XX
XX
Sequence 8, Application US/08304626
Patent No. 5616495
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meyer, Henry
APPLICANT: Muck, Heidi Jane
APPLICANT: Fongcrada, Luis
APPLICANT: Schwab, George E.
TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates
TITLE OF INVENTION: Active Against Hymenoptera Pests and Genes Encoding
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Salimanthik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US-08-304,626
FILING DATE:
CLASSIFICATION: 455
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/987,980
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Salimanthik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/S-104
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO.: A-
SEQUENCE CHARACTERISTICS:
LENGTH: 1245 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS8603
IMMEDIATE SOURCE:
LIBRARY: LAMBDAEM (tm) - 11 library
CLONE: 8603A
SEQUENCE: 1245 AA: 139775 MW: 7963406 CN:
Query Match 11.8% Score 75; DB 1; Length 1245;
Best Local Similarity 28.6% Pred. NO. 7,96e+01;
Matches 10; Conservative 11; Mismatches 13; Indels 1; Gaps 1;
DB 188 VDSIVYSPKELGGLVAVYINSNNRHVGLRSL 422
UY 46 VNNI-VTANNPEETILEYVIVENHPVILP 69
RESULT 14
ID US-08-611-928-8 STANDARD; PRI: 1245 AA.

XX
AC
XX
DT
XX
Sequence 8, Application US/08304626
Patent No. 5824792
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meyer, Henry
APPLICANT: Muck, Heidi Jane
APPLICANT: Fongcrada, Luis
APPLICANT: Schwab, George E.
TITLE OF INVENTION: No. 5824792el Bacillus thuringiensis Isolates
TITLE OF INVENTION: Active Against Hymenoptera Pests
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Salimanthik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/11,928
FILING DATE: 24 MAY 1994
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/156,242
FILING DATE: 24 MAY 1994
APPLICATION NUMBER: US/07/987,980
FILING DATE: 22 MAY 1992
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/797,645
FILING DATE: 25 NOV 1991
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/793,977
FILING DATE: 22 MAY 1991
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Salimanthik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/S-104
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO.: A-
SEQUENCE CHARACTERISTICS:
LENGTH: 1245 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: BACILLUS THURINGIENSIS
INDIVIDUAL ISOLATE: PS8603
IMMEDIATE SOURCE:
LIBRARY: LAMBDAEM (tm) - 11 library

WISER

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, T.K.
Distribution rights by Oxford Molecular Ltd

 Msrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Aug 25 17:51:48 2000; Maspar time 4.43 seconds
 Tabular output not generated. 487,099 Million cell updates/sec

Title: >US-09-257-585-4
 Description: (1-89) from US09257585.pep
 Perfect Score: 634
 Sequence: 1 MKVQATFATLIALAFPPA.....DRVLSGDEFTSAYADDESC 89

Scoring table:
 PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a:geneseq16
 1:geneseqp

Statistics: Mean 28.189; Variance 114.078; scale 0.247

Prod. No. is the number of results produced by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Prod. No.
1	86	13.6	396	1	W23327	Morifilius giganteris xy	1.86e+01
2	82	12.9	315	1	W24867	Human secreted protein	4.38e+01
3	81	12.8	713	1	W05181	Neisseria gonorrhoeae	4.58e+01
4	81	12.8	1394	1	B38634	S-py-vp55 Yp virus gp	4.58e+01
5	81	12.8	1483	1	P80474	Sequence of protein of	4.58e+01
6	79	12.5	412	1	P16647	Syn341 polybasic	5.52e+01
7	79	12.5	713	1	W05180	Neisseria gonorrhoeae	6.52e+01
8	79	12.5	1094	1	W29683	Human Np-K-201 cofactor	6.52e+01
9	78	12.4	286	1	W23326	C protein coupled odor	7.78e+01
10	78	12.4	286	1	K48744	C protein coupled odor	7.78e+01
11	77	12.1	119	1	W98974	Alcaligenes sp. proteo	9.28e+01
12	76	12.0	214	1	K25794	M-2SF isolated from hu	1.10e+02
13	76	12.0	306	1	K63791	Aspergillus nidulans	1.10e+02
14	76	12.0	461	1	W24456	Aspergillus nidulans	1.10e+02
15	76	12.0	540	1	M14772	Atmosin/V chelator tox	1.10e+02
16	76	12.0	772	1	M14438	Type I carboxylate pape	1.10e+02
17	75	11.8	227	1	K25421	Humicola xylinase	1.31e+02
18	75	11.8	227	1	K25421	Humicola xylinase	1.31e+02
19	75	11.8	493	1	W05496	Human TIE ligand M1 P	1.31e+02
20	75	11.8	493	1	W05496	Human TIE ligand M1 P	1.31e+02
21	75	11.8	495	1	W88482	Human papillomavirus	1.31e+02
22	75	11.8	903	1	K27650	Sequence of bovine pap	1.31e+02
23	75	11.8	1245	1	K29029	Human calcium channel	1.31e+02
						Bacillus thuringiensis	1.31e+02

24	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
25	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
26	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
27	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
28	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
29	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
30	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
31	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
32	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
33	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
34	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
35	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
36	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
37	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
38	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
39	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
40	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
41	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
42	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
43	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
44	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02
45	75	11.8	1245	1	W23327	Bacillus thuringiensis	1.31e+02

ALL COMMENTS

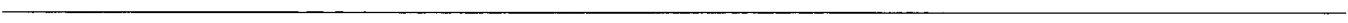
RESULT 1
 ID W23327 standard; length 396 AA.
 AC W23327;
 DE 14-JAN-1998 (first entry)
 DE Morifilius giganteris xylinase.
 KW Morifilius giganteris xylinase.
 KW agricultural waste; fiber production; paper ble
 KW filamentous fungi; alcohol cell wall degradati
 OS Morifilius giganteris CBS 621.95.
 PN W09727290-A1.
 PD 31-JUL-1997.
 PE 22-JAN-1997; DK00480.
 PR 22-JAN-1996; DK00064.
 FA (NVO) NVO-NORDISK AS.
 PI Andersen LN, Dydahl LO, Knappegen MS, Kotof LV
 DP W01: 97-393678/39.
 DR N-PSDB: 164848.
 PT Morifilius giganteris CBS 621.95. DNA encoding a x
 PT providing enzyme for food preparation, animal fo
 PT production and paper bleaching.
 PS Claim 14: Pages 49-49; 1497; English.
 CC This xylinase polypeptide sequence of 396 amino
 CC by a cDNA sequence from Morifilius giganteris str
 CC the xylinase encoding part of the DNA sequence
 CC plasmid pYES 2.0 present in E. coli DSM 5901.
 CC techniques. Host cells transformed with xylinase
 CC xylinases have found cell wall degrading activity
 CC catalyze degradation of the lactone or side of
 CC polysaccharide. The present xylinase enzyme can
 CC inclusion in compositions, optionally with other
 CC for degrading/modifying plant cell walls. The x
 CC are useful in food preparation, production of a
 CC products, separation of various (e.g. wheat int
 CC wine or fruit juice, production or as additives
 CC process animal feed. They can be used in food
 CC animal feed, and to reduce the viscosity of fi
 CC material. They are also useful in pulp or pape
 CC pulp bio-bleaching and in enzymatic breakdown
 CC wastes to produce alcohol fuels.
 SQ Sequence: 396 AA;
 Query Match: 13.6%; Score 86; ID: 1;
 Post local similarity: 4.9%; Prod. No. 1.31e+02
 Matches: 11; (false positive): 7; Mismatches: 28

Sun Aug 27 16:02:00 2000

US-09-257-585-4.rag

Page

Job time : 21 secs.





KW herpes simplex virus (CP4; HSV CP4; modulator; apoptosis)
 KM stimulation; inhibition; HSV infection.
 ON herpes simplex virus.
 PN W09846647-A2.
 PD 22-OCT-1998.
 PE 16-APR-1998; 007571.
 PK 16-APR-1997; US-843654.
 PA (ARCH) ARCH DEV CORP.
 PI Leopoldi R, Kotzman B.
 PR W01: 98-594559/50.
 PS N-1508; V68520.
 PT Use of herpes simplex virus (HS) polypeptide - for developing
 ET products for modulating apoptosis in cells and for identifying
 PT compounds which act as stimulators or inhibitors of apoptosis
 PS Example 2: Pages 63-67; 85pp; English.
 CC This is the amino acid sequence of Herpes simplex virus CP4 used in
 CC the method of the invention as modulators of apoptosis. The methods
 CC and products can be used to identify compounds which modulate
 CC (stimulate or inhibit) apoptosis in cells. They can be used to
 CC immortalise cells for the study of these cells or for growing cells
 CC in large numbers for the production of proteins. They can also be
 CC used for stimulating apoptosis in cells, e.g. for treating a subject
 CC with a HSV infection.
 SN Sequence 1298 AA.

Query Match: 13.4%; Score 82; DB 1; Length 1768.
 Best Local Similarity 28.4%; Pred. No. 1,776-01.
 Matches 1%; Conservative 1%; Mismatches 2%; Indels 4; Gaps 4;

DB 1000 11 STRAFANAVETGATASAPRPIVNT VPAQMP-ADGPAVSPQIYACELLP 1057
 11
 7 LAMTQF-VANARSTSLPFLFKILLATFTSTIFAHNSPVYF-IPYPPGLLS 64

DB 1058 AVCAVR 1064
 11
 65 SSGCLER 71

RESULT 12
 ID W00314 standard; Protein; 810 AA.
 AC W00314:
 DT 01-FEB-1999 (first entry)
 DE CIIIA (class II transactivator) type IV protein.
 KW CIIIA (class II transactivator) type IV protein; MHC class II molecule;
 KW Interferon-gamma; Interleukin-4; Vaccine; Cancer treatment.
 OS Homo sapiens.
 PN EP-87404-A1.
 PD 28-OCT-1998.
 PE 21-APR-1998; 400968.
 PR 22-APR-1997; EP-004954.
 PA (TRC) TRANSCHE SA.
 PI March B.
 PS W01: 98-559115/48.
 PT DNA encoding MHC class II trans-activator polypeptide(s) - useful
 PT e.g. as primers for enzymatic amplification, as detection probes or
 PT as inhibitors of expression of genes encoding MHC in dendritic cells
 PS Disclosure: Pages 66-69; 86pp; French.
 CC The present sequence represents type IV CIIIA (class II transactivator)
 CC protein. The products can be used to treat diseases for which
 CC enhanced expression of genes coding for MHC class II molecules is
 CC desired, especially where the enhanced expression is desired in
 CC dendritic cells or after induction by a cytokine, especially
 CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
 CC can be used to treat diseases for which reduced expression of genes
 CC coding for MHC class II molecules is desired or can be used as vaccines,
 CC especially for cancer treatment.
 SN Sequence 830 AA.

Query Match: 13.98%; Score 79; DB 1; Length 800.
 Best Local Similarity 29.38%; Pred. No. 3,196-01.
 Matches 1%; Conservative 1%; Mismatches 11; Indels 2; Gaps 2;

DB 473 LUPSAASVTPKSVIAYLP-LQF-SILPAGLLEILH 511

ID W57057 standard; Protein; 100 AA.
 AC W57057:
 DT 17-APR-1998 (first entry)
 DE Class II trans activator (CIIIA) 151 deletion anti-
 KW Class II trans activator; MHC; CIIIA; autoimmune;
 KW Transplantation; xenograft; anti; histocompatibility;
 KW gene therapy; arthritis; rheumatoid; 151 deletion mo-
 OS Homo sapiens.
 PN Synuclein.
 PS N-1508; V68520.
 PT Use of herpes simplex virus (HS) polypeptide - for developing
 ET products for modulating apoptosis in cells and for identifying
 PT compounds which act as stimulators or inhibitors of apoptosis
 PS Example 2: Pages 63-67; 85pp; English.
 CC This is the amino acid sequence of Herpes simplex virus CP4 used in
 CC the method of the invention as modulators of apoptosis. The methods
 CC and products can be used to identify compounds which modulate
 CC (stimulate or inhibit) apoptosis in cells. They can be used to
 CC immortalise cells for the study of these cells or for growing cells
 CC in large numbers for the production of proteins. They can also be
 CC used for stimulating apoptosis in cells, e.g. for treating a subject
 CC with a HSV infection.
 SN Sequence 1298 AA.

Query Match: 13.98%; Score 79; DB 1; Length 800.
 Best Local Similarity 29.38%; Pred. No. 3,196-01.
 Matches 1%; Conservative 1%; Mismatches 11; Indels 2; Gaps 2;

DB 1000 11 STRAFANAVETGATASAPRPIVNT VPAQMP-ADGPAVSPQIYACELLP 1057
 11
 7 LAMTQF-VANARSTSLPFLFKILLATFTSTIFAHNSPVYF-IPYPPGLLS 64

DB 1058 AVCAVR 1064
 11
 65 SSGCLER 71

RESULT 14
 ID W00314 standard; Protein; 810 AA.
 AC W00314:
 DT 01-FEB-1999 (first entry)
 DE CIIIA (class II transactivator) type IV protein.
 KW CIIIA (class II transactivator) type IV protein; MHC class II molecule;
 KW Interferon-gamma; Interleukin-4; Vaccine; Cancer treatment.
 OS Homo sapiens.
 PN EP-87404-A1.
 PD 28-OCT-1998.
 PE 21-APR-1998; 400968.
 PR 22-APR-1997; EP-004954.
 PA (TRC) TRANSCHE SA.
 PI March B.
 PS W01: 98-559115/48.
 PT DNA encoding MHC class II trans-activator polypeptide(s) - useful
 PT e.g. as primers for enzymatic amplification, as detection probes or
 PT as inhibitors of expression of genes encoding MHC in dendritic cells
 PS Disclosure: Pages 66-69; 86pp; French.
 CC The present sequence represents type IV CIIIA (class II transactivator)
 CC protein. The products can be used to treat diseases for which
 CC enhanced expression of genes coding for MHC class II molecules is
 CC desired, especially where the enhanced expression is desired in
 CC dendritic cells or after induction by a cytokine, especially
 CC interferon-gamma or interleukin-4. Inhibitors of the gene or protein
 CC can be used to treat diseases for which reduced expression of genes
 CC coding for MHC class II molecules is desired or can be used as vaccines,
 CC especially for cancer treatment.
 SN Sequence 830 AA.

Query Match: 13.98%; Score 79; DB 1; Length 800.
 Best Local Similarity 29.38%; Pred. No. 3,196-01.
 Matches 1%; Conservative 1%; Mismatches 11; Indels 2; Gaps 2;

DB 473 LUPSAASVTPKSVIAYLP-LQF-SILPAGLLEILH 511

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: MS-DOS/MS-DOS
 SOFTWARE: Patcutil release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/0604 474
 FILING DATE: 26-JUL-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Hildander, Steven L.
 REGISTRATION NUMBER: 37,642
 REFERENCE/DOCKET NUMBER: ARIID-239
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 512/418-3000
 TELEFAX: 512/474-7572
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1298 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 SEQ ID NO: 1298 AA: 129493 WW: 7197954 CR:
 Query Match 14.44: Score 82; FR 2; Length 1298;
 Best Local Similarity 28.44; Pred. No. 1.28e-01;
 Matches 19; Conservative 19; Mismatches 25; Indels 4; Gaps 4;
 DB 1000 LISTED: AFGAGVPHGCIASADHRI IVNI-VPAQWMP-ARQYAVSGYAVACHP 1057
 Y 7 LMTQGF-PNAPRSTSPETLFFNITLAVETSTFEAPRNSVYLDGTHYDGLS 64
 LB 1058 AVQAVR 1064
 Y 65 SSGCTER 71
 RESULT 5
 ID US-08-519-6A7A-6 STANDARD: PRI: 1130 AA.
 XX
 AC xxxxxx
 DE
 XX
 DI
 XX
 DE
 XX
 Sequence 6, Application US/08519647A
 Patent No. 5994082
 GENERAL INFORMATION:
 APPLICANT:
 TITLE OF INVENTION: Proteins Essential for the Expression of
 TITLE OF INVENTION: Vertebrate MG Class II Dens. JNA Sequences Encoding Same
 TITLE OF INVENTION: and Pharmacological Compositions
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FISH & NEAVE
 STREET: 1261 AVENUE OF THE AMERICAS
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: U.S.A.
 ZIP: 10020-1104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WordPerfect 6.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/0719 647A
 FILING DATE: 25-AUG-1995
 CLASSIFICATION: 435
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 26 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 SEQ ID NO: 26 AA: 129493 WW: 7197954 CR:
 Query Match 14.44: Score 82; FR 2; Length 1298;
 Best Local Similarity 28.44; Pred. No. 1.28e-01;
 Matches 19; Conservative 19; Mismatches 25; Indels 4; Gaps 4;
 DB 1000 LISTED: AFGAGVPHGCIASADHRI IVNI-VPAQWMP-ARQYAVSGYAVACHP 1057
 Y 7 LMTQGF-PNAPRSTSPETLFFNITLAVETSTFEAPRNSVYLDGTHYDGLS 64
 LB 1058 AVQAVR 1064
 Y 65 SSGCTER 71
 RESULT 5
 ID US-08-519-6A7A-6 STANDARD: PRI: 1130 AA.
 XX
 AC xxxxxx
 DE
 XX
 DI
 XX
 DE
 XX
 Sequence 6, Application US/08519647A
 Patent No. 5994082
 GENERAL INFORMATION:
 APPLICANT:
 TITLE OF INVENTION: Proteins Essential for the Expression of
 TITLE OF INVENTION: Vertebrate MG Class II Dens. JNA Sequences Encoding Same
 TITLE OF INVENTION: and Pharmacological Compositions
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FISH & NEAVE
 STREET: 1261 AVENUE OF THE AMERICAS
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: U.S.A.
 ZIP: 10020-1104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WordPerfect 6.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: EP9411378.7
 FILING DATE: 26-AUG-1994

ATTORNEY/AGENT INFORMATION:
 NAME: HALEY, JAMES E.
 REGISTRATION NUMBER: 27,704
 REFERENCE/DOCKET NUMBER: US-11
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-596-9000
 TELEFAX: 212-596-9000
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1140 amino acids
 TYPE: amino acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: RNA (mRNA)
 HYDROLICALLY: NO
 AMI-SUM: NO
 SEQ ID NO: 1130 AA: 124477 WW: 6274599 CR:
 Query Match 14.94: Score 79; DB 2; Length
 Best Local Similarity 29.84; Pred. No. 2.18e-01;
 Matches 12; Conservative 14; Mismatches 11;
 DB 723 LISTED: AAGAGVPHGCIASADHRI IVNI-VPAQWMP-ARQYAVSGYAVACHP 1057
 Y 7 LMTQGF-PNAPRSTSPETLFFNITLAVETSTFEAPRNSVYLDGTHYDGLS 64
 LB 1058 AVQAVR 1064
 Y 65 SSGCTER 71
 RESULT 6
 ID US-08-519-6A7A-6 STANDARD: PRI: 213 AA
 XX
 AC xxxxxx
 DE
 XX
 DI
 XX
 DE
 XX
 Sequence 48, Application US/08519647A
 Patent No. 584084
 GENERAL INFORMATION:
 APPLICANT: HARKO, Joseph
 APPLICANT: PHOENIX, Matthew R.
 APPLICANT: NO, Michael J.
 APPLICANT: ABULIN, Boris I.
 APPLICANT: KODOL, Andrew S.
 TITLE OF INVENTION: Proteins Inhibiting the
 TITLE OF INVENTION: of 121 kDa
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: HARKO and WINTER, P.O.
 STREET: 5476 Mulberry Street, Suite 200
 CITY: Houston
 STATE: Texas
 COUNTRY: US
 ZIP: 77034
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: Patcutil release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/0604 474
 FILING DATE: 26-JUL-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Hildander, Steven L.
 REGISTRATION NUMBER: 37,642
 REFERENCE/DOCKET NUMBER: ARIID-239
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 512/418-3000
 TELEFAX: 512/474-7572
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 264 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 SEQ ID NO: 264 AA: 129493 WW: 7197954 CR:

STRANDLINESS: S
TOPOLOGY: linear

OTHER INFORMATION:	nucleotides 1 through 102 of sequence ID No. 5852171
Patent No.	5852171

00	REFERENCE/SEQUENCE NUMBER:	284268
01	TELEPHONICAL INFORMATION:	
02	TELEPHONE: (404) 874-4794	
03	FILMAY: (404) 874-4795	
04	INFORMATION FOR SEQ. ID NO. 1:	
05	SEQUENCE CHARACTERISTICS:	
06	LENGTH: 248 amino acids	
07	TYPE: amino acid	
08	TOXICITY: 1 (low)	
09	MOLECULE TYPE: protein	
10	HYDROPHILIC: N	
11	FEATURE:	
12	NAME/KEY: msc284268.1	
13	LOCATION: 1-248	
14	OTHER INFORMATION: /note="Proteinase 2"	
15	OTHER INFORMATION: accepted conserved by motif	
16	Patent No. 6,037,456	
17	OTHER INFORMATION: N. 6,037,450 1."	
18	FEATURE:	
19	NAME/KEY: Msc284268.1	
20	LOCATION: 1-248	
21	OTHER INFORMATION: /note="Amino acids 1-248"	
22	OTHER INFORMATION: /note="Amino acids 1-248"	
23	OTHER INFORMATION: /note="Amino acids 1-248"	
24	FEATURE:	
25	NAME/KEY: Msc284268.1	
26	LOCATION: 1-248	
27	OTHER INFORMATION: /note="Amino acids 1-248"	
28	OTHER INFORMATION: /note="Amino acids 1-248"	
29	OTHER INFORMATION: /note="Amino acids 1-248"	
30	FEATURE:	
31	NAME/KEY: Msc284268.1	
32	LOCATION: 1-248	
33	OTHER INFORMATION: /note="Amino acids 1-248"	
34	OTHER INFORMATION: /note="Amino acids 1-248"	
35	OTHER INFORMATION: /note="Amino acids 1-248"	
36	FEATURE:	
37	NAME/KEY: Msc284268.1	
38	LOCATION: 1-248	
39	OTHER INFORMATION: /note="Amino acids 1-248"	
40	OTHER INFORMATION: /note="Amino acids 1-248"	
41	OTHER INFORMATION: /note="Amino acids 1-248"	
42	FEATURE:	
43	NAME/KEY: Msc284268.1	
44	LOCATION: 1-248	
45	OTHER INFORMATION: /note="Amino acids 1-248"	
46	OTHER INFORMATION: /note="Amino acids 1-248"	
47	OTHER INFORMATION: /note="Amino acids 1-248"	
48	FEATURE:	
49	NAME/KEY: Msc284268.1	
50	LOCATION: 1-248	
51	OTHER INFORMATION: /note="Amino acids 1-248"	
52	OTHER INFORMATION: /note="Amino acids 1-248"	
53	OTHER INFORMATION: /note="Amino acids 1-248"	
54	FEATURE:	
55	NAME/KEY: Msc284268.1	
56	LOCATION: 1-248	
57	OTHER INFORMATION: /note="Amino acids 1-248"	
58	OTHER INFORMATION: /note="Amino acids 1-248"	
59	OTHER INFORMATION: /note="Amino acids 1-248"	
60	FEATURE:	
61	NAME/KEY: Msc284268.1	
62	LOCATION: 1-248	
63	OTHER INFORMATION: /note="Amino acids 1-248"	
64	OTHER INFORMATION: /note="Amino acids 1-248"	
65	OTHER INFORMATION: /note="Amino acids 1-248"	
66	FEATURE:	
67	NAME/KEY: Msc284268.1	
68	LOCATION: 1-248	
69	OTHER INFORMATION: /note="Amino acids 1-248"	
70	OTHER INFORMATION: /note="Amino acids 1-248"	
71	OTHER INFORMATION: /note="Amino acids 1-248"	
72	FEATURE:	
73	NAME/KEY: Msc284268.1	
74	LOCATION: 1-248	
75	OTHER INFORMATION: /note="Amino acids 1-248"	
76	OTHER INFORMATION: /note="Amino acids 1-248"	
77	OTHER INFORMATION: /note="Amino acids 1-248"	
78	FEATURE:	
79	NAME/KEY: Msc284268.1	
80	LOCATION: 1-248	
81	OTHER INFORMATION: /note="Amino acids 1-248"	
82	OTHER INFORMATION: /note="Amino acids 1-248"	
83	OTHER INFORMATION: /note="Amino acids 1-248"	
84	FEATURE:	
85	NAME/KEY: Msc284268.1	
86	LOCATION: 1-248	
87	OTHER INFORMATION: /note="Amino acids 1-248"	
88	OTHER INFORMATION: /note="Amino acids 1-248"	
89	OTHER INFORMATION: /note="Amino acids 1-248"	
90	FEATURE:	
91	NAME/KEY: Msc284268.1	
92	LOCATION: 1-248	
93	OTHER INFORMATION: /note="Amino acids 1-248"	
94	OTHER INFORMATION: /note="Amino acids 1-248"	
95	OTHER INFORMATION: /note="Amino acids 1-248"	
96	FEATURE:	
97	NAME/KEY: Msc284268.1	
98	LOCATION: 1-248	
99	OTHER INFORMATION: /note="Amino acids 1-248"	
100	OTHER INFORMATION: /note="Amino acids 1-248"	
101	OTHER INFORMATION: /note="Amino acids 1-248"	
102	FEATURE:	
103	NAME/KEY: Msc284268.1	
104	LOCATION: 1-248	
105	OTHER INFORMATION: /note="Amino acids 1-248"	
106	OTHER INFORMATION: /note="Amino acids 1-248"	
107	OTHER INFORMATION: /note="Amino acids 1-248"	
108	FEATURE:	
109	NAME/KEY: Msc284268.1	
110	LOCATION: 1-248	
111	OTHER INFORMATION: /note="Amino acids 1-248"	
112	OTHER INFORMATION: /note="Amino acids 1-248"	
113	OTHER INFORMATION: /note="Amino acids 1-248"	
114	FEATURE:	
115	NAME/KEY: Msc284268.1	
116	LOCATION: 1-248	
117	OTHER INFORMATION: /note="Amino acids 1-248"	
118	OTHER INFORMATION: /note="Amino acids 1-248"	
119	OTHER INFORMATION: /note="Amino acids 1-248"	
120	FEATURE:	
121	NAME/KEY: Msc284268.1	
122	LOCATION: 1-248	
1		

00 APPLICANT: Esmen, Charles T.
 01 TITLE OF INVENTION: Cloning and Regulation of an Endothelial
 02 CELL PROTEIN C/ACTIVATED PROTEIN C RECEPTOR
 03 NUMBER OF SEQUENCES: 6
 04 CORRESPONDENCE ADDRESS:
 05 ADDRESSEE: Patricia L. Pabst
 06 STREET: 2800 One Atlantic Center, 1201 West Peachtree
 07 STREET: Street
 08 CITY: Atlanta
 09 STATE: Georgia
 10 COUNTRY: US
 11 ZIP: 30306-4450
 12 COMPUTER READABLE FORM:
 13 MEDIUM TYPE: Floppy disk
 14 COMPUTER: IBM PC compatible
 15 OPERATING SYSTEM: PC-DOS/MS-DOS
 16 SOFTWARE: Patent in Release #1.0, Version #1.25
 17 CURRENT APPLICATION DATA:
 18 APPLICATION NUMBER: US/08/200-400A
 19 FILING DATE: 12-AUG-1994
 20 CLASSIFICATION: 435
 21 ATTORNEY/AGENT INFORMATION:
 22 NAME: Pabst, Patricia L.
 23 REGISTRATION NUMBER: 31,284
 24 REFERENCE/BOOKET NUMBER: OMRI52
 25 TELECOMMUNICATION INFORMATION:
 26 TELEPHONE: (404)873-8794
 27 TELEFAX: (404)873-8795
 28 INFORMATION FOR SEQ ID NO: 2:
 29 SEQUENCE CHARACTERISTICS:
 30 LENGTH: 238 amino acids
 31 TYPE: amino acid
 32 TOPOLOGY: linear
 33 MOLECULE TYPE: protein
 34 HYPOTHEICAL: NO
 35 FEATURE:
 36 NAME/KEY: misc_feature
 37 LOCATION: 1..365
 38 OTHER INFORMATION: /note= "Prothelial Cell Protein Receptor"
 39 OTHER INFORMATION: encoded by
 40 OTHER INFORMATION: nucleotides 1 through 1302 of Sequence ID No. 5695993.1.
 41 Patent No. 5695993
 42 FEATURE:
 43 NAME/KEY: Modified-site
 44 LOCATION: 1..15
 45 OTHER INFORMATION: /note= "Amino acids 1-15 represent
 46 OTHER INFORMATION: a putative signal sequence."
 47 FEATURE:
 48 NAME/KEY: Domain
 49 LOCATION: 211..236
 50 OTHER INFORMATION: /note= "Amino acids 211-236
 51 OTHER INFORMATION: represent a putative transmembrane domain."
 52 FEATURE:
 53 NAME/KEY: Active-site
 54 LOCATION: 47..174
 55 OTHER INFORMATION: /note= "Amino acids 47-49, 64-66,
 56 OTHER INFORMATION: 136-138 and 172-174 represent potential
 57 OTHER INFORMATION: N-glycosylation sites."
 58 FEATURE:
 59 NAME/KEY: Modified-site
 60 LOCATION: 17..186
 61 OTHER INFORMATION: /note= "Amino acids 17, 114, 118
 62 OTHER INFORMATION: and 186 represent extracellular cysteine
 63 OTHER INFORMATION: residues."
 64 SEQUENCE: 238 AA; 26641 MW; 404069 CN;
 65
 66 Query Match 13.7% Score 78; DB 1; Length 238;
 67 Best Local Similarity 33.3% Pred. No. 2,63e+01;
 68 Matches 13; Conservative 11; Mismatches 11; Indels 1, Gaps 1.
 69
 70 3 TTLLPILLSSWATSSQASWNGQIMHGLSYRNYH 41
 71 121 1111 1 111 111 111 111
 72 21 TSLPFLFKNLIAVETISTLEAVINSESYRIHYF 58

00 RESULT 11
 01 ID US-08-884-203-2 STANDARD: PRT: 238 AA.
 02 XX
 03 XXXXXX
 04
 05 Sequence 2, Application US/08884203
 06
 07 Patent No. 5804392
 08 GENERAL INFORMATION:
 09 APPLICANT: Charles T. Esmen, Deborah J. Stearns-Kurosawa and
 10 APPLICANT: Shinichiro Kurosawa
 11 TITLE OF INVENTION: DIAGNOSTIC ASSAYS USING SOLUBLE ENDOTHELIAL
 12 CELL PROTEIN C/ACTIVATED PROTEIN C RECEPTOR
 13 NUMBER OF SEQUENCES: 8
 14 CORRESPONDENCE ADDRESS:
 15 ADDRESSEE: Patricia L. Pabst
 16 STREET: 2800 One Atlantic Center
 17 STREET: 1201 West Peachtree Street
 18 CITY: Atlanta
 19 STATE: Georgia
 20 COUNTRY: USA
 21 ZIP: 30309-3450
 22 COMPUTER READABLE FORM:
 23 MEDIUM TYPE: Floppy disk
 24 COMPUTER: IBM PC compatible
 25 OPERATING SYSTEM: PC-DOS/MS-DOS
 26 SOFTWARE: Patent in Release #1.0, Version #1.25
 27 CURRENT APPLICATION DATA:
 28 APPLICATION NUMBER: US/08/884,203
 29 FILING DATE:
 30 CLASSIFICATION: 435
 31 ATTORNEY/AGENT INFORMATION:
 32 NAME: Pabst, Patricia L.
 33 REGISTRATION NUMBER: 31,284
 34 REFERENCE/BOOKET NUMBER: OMRI168
 35 TELECOMMUNICATION INFORMATION:
 36 TELEPHONE: (404) 873-8794
 37 TELEFAX: (404) 873-8795
 38 INFORMATION FOR SEQ ID NO: 2:
 39 SEQUENCE CHARACTERISTICS:
 40 LENGTH: 238 amino acids
 41 TYPE: amino acid
 42 TOPOLOGY: linear
 43 MOLECULE TYPE: protein
 44 HYPOTHEICAL: NO
 45 FEATURE:
 46 NAME/KEY: misc_feature
 47 LOCATION: 1..365
 48 OTHER INFORMATION: /note= "Endothelial Cell Protein
 49 OTHER INFORMATION: Receptor encoded by nucleotides 1 through 1302 of Sequn
 50 Patent No. 5804392
 51 OTHER INFORMATION: No. 5801392.1."
 52 FEATURE:
 53 NAME/KEY: Modified-site
 54 LOCATION: 1..15
 55 OTHER INFORMATION: /note= "Amino acids 1-15 represent
 56 OTHER INFORMATION: a putative signal sequence."
 57 FEATURE:
 58 NAME/KEY: Domain
 59 LOCATION: 211..236
 60 OTHER INFORMATION: /note= "Amino acids 211-236
 61 OTHER INFORMATION: represent a putative transmembrane domain."
 62 FEATURE:
 63 NAME/KEY: Active-site
 64 LOCATION: 47..174
 65 OTHER INFORMATION: /note= "Amino acids 47-49, 64-66,
 66 OTHER INFORMATION: 136-138 and 172-174 represent potential
 67 OTHER INFORMATION: N-glycosylation sites."
 68

[illegible]

TELECOMMUNICATION INFORMATION:
TELEPHONE: 619 546 1555
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 141 amino acids
TYPE: AMINO ACID
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
FEATURE:
NAME/KEY: Region
LOCATION: 1..141
OTHER INFORMATION: "The 141 amino acid sequence predicted from the nucleic acid product which results from amplification of the mouse Alpha 6b OTHER INFORMATION: cDNA with primers 1157/1156."
FEATURE:
NAME/KEY: Domain
LOCATION: 88..113
OTHER INFORMATION: "The putative transmembrane FEATURE:
OTHER INFORMATION: domain."
NAME/KEY: Region
LOCATION: 1..120
OTHER INFORMATION: "Seq ID NO:5 is identical to OTHER INFORMATION: SEQ ID NO:7 at amino acid position 1 through 120; OTHER INFORMATION: the two sequences diverge at amino acid 121."
SEQUENCE 14 AA: 16105 MW: 103520 CN:
Query Match 13.4% Score 76, DB 1, Length 141;
Best Local Similarity 43.5% Pred. No. 3,836+01;
Matches 19, Conservative 1, Mismatches 9, Gaps 1,
Db 109 VF-ILMKGEFKRSYRDSIFRY 130
11 111 11 111 111
4 VPSITWTFQFANAPQSTIVRF 26
RESULT 14
ID US-08-460-907B-2 STANDARD: PRT: 1531 AA.
AC xxxxxx
XX
XX
XX
XX
Sequence 2, Application US/08460907B
Patent No. 5891724
GENERAL INFORMATION:
APPLICANT: Deeley, Roger G.
APPLICANT: Cole, Susan P.C.
TITLE OF INVENTION: METHODS FOR CONFERING MULTIDRUG
TITLE OF INVENTION: RESISTANCE ON A CELL.
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
STREET: Queen's University at Kingston
CITY: Kingston
STATE: Ontario
COUNTRY: CANADA
ZIP: K7L 3N6
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,907B
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/906,923
FILING DATE: 27-OCT-1992
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/029,340
FILING DATE: 8-MAR-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/141,893
FILING DATE: 26-OCT-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/407,207
FILING DATE: 20-MAR-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Steeg, Carol Mironicki
REGISTRATION NUMBER: 39,539
REFERENCE/DOCKET NUMBER: 01551
TELECOMMUNICATION INFORMATION:
TELEPHONE: (613) 545 2342
TELEFAX: (613) 545-6853
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1531 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 1531 AA: 17671 MW: 1235123 CN:
Query Match 13.4% Score 76, DB 2, Length 1531;
Best Local Similarity 27.3% Pred. No. 3,836+01;
Matches 15, Conservative 14, Mismatches 24, Gaps 2,
Db 1435 KQIVCIARALKIKILYDAIAVDIETDILQSTIRIQFEDCVIITAHMLN 1489
11 111 11 111 111 111 111 111 111 111 111 111 111 111 111
18 KQSTIRIRLKHNLIAAE-ISHKANE-SPSYIKHYKDCGLSSQCLF 70
RESULT 15
ID US-08-461-384B-2 STANDARD: PRT: 1531 AA.
AC xxxxxx
XX
XX
XX
XX
XX
Sequence 2, Application US/08461384B
Patent No. 6025473
GENERAL INFORMATION:
APPLICANT: Cole, Susan P.C.
APPLICANT: Deeley, Roger G.
TITLE OF INVENTION: MULTIDRUG RESISTANCE PROTEINS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: PARTEQ RESEARCH & DEVELOPMENT INNOVATIONS
STREET: Queen's University at Kingston
CITY: Kingston
STATE: Ontario
COUNTRY: CANADA
ZIP: K7L 3N6
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,384B
FILING DATE: 05-JUN-95
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/906,923

00 FILING DATE: 27-OCT-1992
 00 APPLICATION NUMBER: 08/029,440
 00 FILING DATE: 8-MAR-1994
 00 APPLICATION NUMBER: 08/141,894
 00 FILING DATE: 26-OCT-1994
 00 APPLICATION NUMBER: 08/407,207
 00 FILING DATE: 20-MAR-1995
 00 ATTORNEY/AGENT INFORMATION:
 00 NAME: Steed, Carol Mernick
 00 REGISTRATION NUMBER: 39,539
 00 REFERENCE/CLIENT NUMBER: 0147
 00 TELECOMMUNICATION INFORMATION:
 00 TELEPHONE: (613) 545-2342
 00 TELEFAX: (613) 545-6853
 00 INFORMATION FOR SEQ ID NO: 2:
 00 SEQUENCE CHARACTERISTICS:
 00 LENGTH: 1541 amino acids
 00 TYPE: amino acid
 00 TOPOLOGY: linear
 00 MOLECULE TYPE: protein
 00 SEQUENCE: 1541 AA; 171671 MW; 1285423 CN;

Query Match 14.4%; Score 76; DB 3; Length 1531;
 Best Local Similarity 27.3%; Pred. No. 3,836,01;
 Matches 15; Conserved: 14; Mismatches 24; Indels 2; Gaps 2;

1435 KQIVTAPATTEKTIIVIPATAVTHFTQULISTPTQPEFCTVLTIAHIN 1489
 00 16 ROSTSLPPELEKLLLAIE-TSTERAHNE-SPSYIRHPRPVQLSSSACLE 70

Search completed: Fri Aug 25 17:51:30 2000
 Job time : 20 secs.



MESSAGE

Release 3.1A John F. Collins, Bioinformatics Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

March_pp protein - protein database search using Smith-Waterman algorithm
Run on: Fri Aug 25 17:48:14 2000; MasPar time 6.61 seconds
549,771 Million cell updates/sec
Tabular output not generated.

Title: >US-09-257-585-3
Description: (1-77) from US09257585.pep
Perfect Score: 568
Sequence: 1 MGVSFLMTLTFANARQS.....KPGGLSSSGLFRLHHPH 77

Scoring table:
PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1664
1:p161 2:p162 3:p163 4:p164

Statistics: Mean 38.246; Variance 70.485; Scale 0.543

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	103	18.1	971	1	A36376	replication licensing
2	92	16.2	1643	2	101359	multidrug-resistance
3	91	16.0	481	2	602688	neurogenic basic helix
4	91	16.0	383	2	104688	neuro D-related factor
5	90	15.8	1861	2	115845	microtubule-associated
6	89	15.7	140	2	115574	hypothetical protein
7	88	15.5	533	2	115116	hypothetical protein
8	87	15.3	231	2	670381	Rhase III; Aquifex a
9	86	15.1	219	2	670381	hypothetical protein
10	84	14.8	489	2	103411	protein RFL - rice
11	83	14.6	157	2	516422	crick protein; Shadab
12	83	14.6	158	2	A57438	tryptophan-rich sero
13	83	14.6	241	2	577512	lung protein; Spwech
14	83	14.6	278	2	549623	crick protein - rhodob
15	82	14.4	278	2	A71216	hypothetical protein
16	82	14.4	240	2	A39842	insulin-like growth I
17	82	14.4	475	2	B73842	carbamoyl phosphate s
18	82	14.4	1298	1	HDB575	immediate-early prote
19	81	14.3	297	2	545747	hypothetical protein
20	81	14.3	256	2	D64942	cobalt transport prot
21	81	14.3	394	2	575004	hypothetical protein
22	80	14.1	515	2	575007	hypothetical protein
23	80	14.1	541	2	A48901	2-hydroxy-3-oxo-5-pyruv

24	80	14.1	637	2	S41506	NADH dehydrog
25	80	14.1	639	2	111607	NADH dehydrog
26	80	14.1	644	1	A40212	uracimidylate
27	80	14.1	644	2	101604	lamin B1
28	80	14.1	671	2	577257	protease-inhib
29	80	14.1	693	1	H07260	DNA-protease
30	80	14.1	760	2	102930	chlorophyll a
31	79	13.9	156	1	E0857	ribosomal
32	79	13.9	442	2	103412	nitrate reduct
33	79	13.9	444	2	S14405	membrane
34	79	13.9	550	2	S09410	ribosomal
35	79	13.9	1100	2	A48843	mito class
36	78	13.7	236	2	A75490	endo-beta
37	78	13.7	373	2	B01520	protease-inhib
38	78	13.7	413	2	101544	hypertonic
39	78	13.7	511	2	103422	2 hydroxy
40	78	13.7	658	1	E0435A	gap peptid
41	78	13.7	699	2	112713	HADH dehydr
42	77	13.6	308	2	516613	opacity p
43	77	13.6	442	2	10760	lactase-like
44	77	13.6	570	2	570642	cytochrome
45	77	13.6	824	2	101642	hypertonic

ALIGNMENTS

RESULT 1
ENTRY 1
TITLE A36376 #type complete
replication licensing factor M36 - y
cerevisiae
ALTERNATE_NAMES
ORGANISM microthromosome mutant virus protein M36
#formal name Saccharomyces cerevisiae
13-Aug-1999 #s1600 revision 1.0
DATE 20-Aug-1999
ACCESSIONS
#authors A36376, Stoeckli, Stoeckli
#journal A36376
#journal A36376
#journal Mol. Cell. Biol. (1999) 19:5707-5720
#title The characterization of the microthromosome as
is characteristic of mutants of yeast
#cross-references M010191042400
#accession A36376
#molecule-type DNA
#residues 1 971 #label C1H
#cross-references S00X54400 NID43910; P16N176
S50491
#authors Dietrich, F.S.
#submission submitted to the EMBL Data Library, to
#description The sequence of the proteinase inhibitor
#accession S50512
#molecule-type DNA
#residues 1 971 #label C1H
#cross-references S00X54400 NID43910; P16N176
S50491
#authors Mulligan, J.L.; Dietrich, F.S.; Jones
Davis, J.W.; Taylor, J.; Walker
#submission submitted to the EMBL Data Library, to
#accession S30851
#molecule-type DNA
#residues 1 971 #label C1H
#cross-references S00X54400 NID43910; P16N176
S50491
#description The complex of six M36 proteins is a
must be found at the origins of replication
DNA replication. As replication proceeds,
phosphorylated and histone from the ori
GENETICS
#name S30851
#cross-references S00X54400 NID43910; P16N176
S50491
#map_position 51
#comment The proteinase inhibitor is a heterodimer
M36X and M36Y. Some times call

[illegible]

```

#accession      G02658
#structure      Preliminary; translated from Chr7Mbl/70MbJ
##molecule_type DNA
##residues      1-381 ##label TAP
##cross_references EMBL:U58601; NID:q1477746; PDB:q1477749
GENETICS
#gene           NeuroD2
SUMMARY         #length 381 #molecular_weight 41300 #checksum 3349

Query Match     16.0% Score 91; DB 2; Length 381,
                Exact Local Similarity 29.7%, Exact No. 2,00E+01;
                Matches 17; Conservation 10; Mismatches 15; Indels 1; Gaps 1;

C1              LLALTEITSTERKANEESTYTHIVIE-KGLSSSSQ 68
          ||| | | | | | | | | | | | | | | | | |
EB 215 FLIPGAMVAGSRHSGCPHAMHPYPSPSLAAGC 251
          ||| | | | | | | | | | | | | | | | | |
C2              LLALTEITSTERKANEESTYTHIVIE-KGLSSSSQ 68

RESULT          4
ENTRY           JC4688
TITLE           neuro D-related factor - mouse
ORGANISM        Mus musculus #common_name house mouse
DATE            10-Mar-1996 #sequence_revision 19-Jul-1996 #text_change
                17 Mar-1999
ACCESSIONS      JC4688
REFERENCE       Yasunami, M.; Suzuki, K.; Maruyama, H.; Kawakami, H.; Natsu,
                Y.; Hagihara, M.; Ohkubo, R.;
                Etzionim, B.; Shigen, K.; Shimizu, A.; Iijima, T.;
                Molecular cloning and characterization of a cDNA encoding a
                novel basic helix-loop-helix protein structurally related
                to Neuro D/bHLH1.
#cross_references MIMD:96183021
#accession      JC4688
##molecule_type mRNA
##residues      1-383 ##label VAS
##cross_references DDBJ:083567; NID:q1304164; PDB:d1012549; PDB:q1304164
#expect_metal_source embryo
COMMENT          This factor plays distinct roles in neural development and
                plasticity as a novel member of the NeuroD family.
FEATURE
16-44          #region proline-rich
83-92          #region glutamic acid-rich
113-127        #region alanine-glycine-rich
SUMMARY         #length 383 #molecular_weight 41494 #checksum 389

Query Match     16.0% Score 91; DB 2; Length 383;
                Exact Local Similarity 29.7%, Exact No. 2,00E+01;
                Matches 17; Conservation 10; Mismatches 15; Indels 1; Gaps 1;

EB 216 FLTPGATCAPEHSQGTFAHHPTFYCSLANAGC 252
          ||| | | | | | | | | | | | | | | | | |
C2              LLALTEITSTERKANEESTYTHIVIE-KGLSSSSQ 68

RESULT          5
ENTRY           TI1845
TITLE           microtubule-associated protein - fruit fly (Drosophila
                melanogaster)
ORGANISM        Drosophila melanogaster
TAG             29-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
                20-Sep-1999
ACCESSIONS      TI1845
REFERENCE       Saunders, R.D.; Andrus, M.C.; Howard, T.; Gonzalez, C.;
                Glover, D.M.;
                J. Cell Biol. (1997) 137:881-890
                The Drosophila gene abnormal spindle encodes a
                microtubule-associated protein that associates with the
                polar regions of the mitotic spindle.
#accession      TI1845
#structure      Preliminary; translated from Chr7Rb/70MbJ
##molecule_type mRNA

```




 RELEASE 3.1A JOHN F. COLLINS, BIOCOMPUTING RESEARCH UNIT,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

 (TM)

 Release 3.1A John F. Collins, BioComputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

Match-EP protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Aug 25 17:48:55 2000; Match time 4.52 seconds
 Labelled output not generated. 528,409 Million cell updates/sec

 Description: MUS-09-257-585-3
 Perfect Score: (1-77) from US09257585 pop
 Sequence: 1 MCVPSLLMTTQTFANAEQS
KPVLLSSQTEELHPTL 77

Scoring table:
 Gap 11
 Gap 11

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: swiss-prot38
 1:swissprot

Statistics: Mean 48.911; Variance 63.226; scale 0.615

Prod. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length DB	ID	Description	Prod. No.
1	103	18.1	971	MCM3_YEAST	4.02e-04
2	96	17.4	374	CAHA_PSEFT	2.32e-03
3	91	16.0	382	NDEF2_HUMAN	4.05e-02
4	91	16.0	383	NDEF2_MOUSE	4.05e-02
5	89	15.7	382	NDEF2_KAI	4.05e-02
6	87	15.3	221	RNC_AQUAE	8.46e-02
7	84	14.8	489	FT_06YSA	1.74e-01
8	84	14.6	162	MEP4_HUMAN	5.06e-01
9	83	14.6	230	C552_HIFE	7.18e-01
10	83	14.6	278	CRIC_RHOSH	7.18e-01
11	82	14.4	240	INSULIN_LIFE	1.02e-03
12	82	14.4	1208	ICG4_HCV1	1.02e-03
13	81	14.3	207	RASHL_CAEEL	1.42e-03
14	81	14.3	758	YLLDM_CAEEL	1.42e-03
15	80	14.1	541	OST_FAT	2.01e-03
16	80	14.1	541	OST_MOUSE	2.01e-03
17	80	14.1	637	N05M_SHIPO	2.01e-03
18	80	14.1	644	UCROM_RAT	2.01e-03
19	80	14.1	690	UCROM_HUMAN	2.01e-03
20	80	14.1	692	UCROM_MOUSE	2.01e-03
21	79	13.9	156	RS7_EYEL	2.01e-03
22	79	13.9	341	IN72_SP101	2.81e-03
23	79	13.9	555	PI04_BRANA	2.81e-03

RESULT ID	1	MCM3_YEAST	STANDARD	443	971 AA
AC	P24279				
DE	01-MAR-1992 (Ref. 21, first sequence update)				
DI	01-MAR-1992 (Ref. 21, last annotation update)				
DI	01-OCT-1996 (Ref. 44, last annotation update)				
DE	MINTCHROM380ME MAINTENANCE (HUMAN)				
GN	MCM3 OR YEL032W OR SYCP1 (Yeast)				
OS	Saccharomyces cerevisiae (Baker's Yeast)				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Sac				
NC	Saccharomycetaceae; Saccharomycetaceae				
LN	11				
RP	SEQUENCE FROM N.A.				
KA	MEHLING, G1042499.				
KA	GILSON S.T., Suroski R., Lee B.K.				
RT	"The phenotype of the mitchromosome maintenance				
RI	characteristic of mutants defective in DNA replication				
RI	Mol. Cell. Biol. 16:3907-3912 (1996).				
LN	12				
RP	SEQUENCE FROM N.A.				
KA	Mulligan J.T., Dietrich F.S., Bonnessy K.M., Sedg				
KA	Mot V., Taylor P., Nakamura K., Roberts L., Davis				
KA	Submitted (May 1997) to GenBank/EMBL data				
LN	13				
RP	SEQUENCE FROM N.A.				
KA	STAIRS-SMITH, A9972.				
KA	Dietrich F.S., Mulligan J.T., Bonnessy K.M., Aylis				
KA	Aylis E., Borno A., Roberts L., Partington J., The				
KA	Young E., Duncan M., Nuzman E., Dietrich F.S., H				
KA	Chiang R., Kayser A., Komp E., Lasker D., Lee H.				
KA	Macdole E., Nakamura K., Smith A., Norrien R.				
KA	Perel F.X., Roberts L., Sedg, P., Sedg, S., Shou				
KA	Taylor P., Mot V., Yellon M., Robinson D., Davis				
LN	14				
RP	FUNCTION: NECESSARY FOR THE PROGRESSION OF				
CC	ROLE IN THE INITIATION OF CHROMOSOMAL DNA REPL				
CC	INTERACTING ACTIVELY WITH ARID-CONTAINING				
CC	(ARF). ITS ARF SPECIFICITY IS NOT IDENTICAL				
CC	MOR2 PROTEIN.				
CC	1. SUBCELLULAR LOCATION: Nucleus.				
CC	2. SIMILARITY: BELONGS TO THE MCM FAMILY.				
CC	3. THIS SWISS-PROT entry is annotated. It is predicted				
CC	between the Swiss Institute of Bioinformatics at				
CC	the European Bioinformatics Institute. There are				
CC	use by non-profit third parties as long as its				
CC	modified and this statement is not removed.				

EMBL: X53540; CAA37616.1; -
 EMBL: U18779; AB65010.1; -
 EMBL: A36376; A36376.
 PIR: S30851; S30851.
 SCQ: L0001039; MCM3.
 PFM: PF00493; MCM_1.
 PROSITE: PS00847; MCM_1; 1.
 PROSITE: PS50051; MCM_2; 1.
 Transcription regulation, DNA binding, Nuclear protein;
 DNA replication; ATP-binding.
 KW DOMAIN 359 566 MCM.
 NP_BIND 409 416 ATP (POTENTIAL).
 SEQUENCE 971 AA; 107517 MW; 43DD4D6C4F4456DC CRC64;

Query Match 18.1%; Score 103; DR 1; Length 971;
 Best Local Similarity 31.7%; Pred. No. 4,026-04;
 Matches 10, Conservative 14, Mismatches 25, Indels 2, Gaps 2.

DB 528 DEPCNIAI PDLISPEFI FVDTDTNEI PPSISEHVPHTHYLPGYI FGEVPEPIN 587
 16 MKQSTSI PPIPIPK-NIIIALEI STIRAHNSPSYIP-HYVPEQGLSSSQI PPIH 73

RESULT 2
 ID GANA_PSEFI STANDARD: PFI: 376 AA.
 AC P48841;
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DI 15-FEB-2000 (Rel. 39, Last annotation update)
 DE APARTINGALACTAN ENDO-1,4-BETA-GALACTOSIDASE PEPPIPSAP (EC 3.2.1.89)
 DE (ENDO-1,4-BETA-GALACTANASE) (GALACTANASE).
 GN GANA OR GANA.
 OS Pseudomonas fluorescens.
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
 CC Pseudomonas.
 RN (1)
 RC SEQUENCE FROM N.A.
 RC STRAIN-SP. CELLULOSA;
 RX MEDLINE: 98060803.
 RA Brattwaite K.L., Barra T., Spurway T.D., Charnock S.J., Black G.W.,
 Hughes N., Lakey J.R., Vidler P., Hazlewood G.J., Heinrich B.,
 Gilbert H.J.;
 *Evidence that galactanase A from Pseudomonas fluorescens subspecies
 PT cellulosa is a retained family 53 glycosyl hydrolase in which E161
 RT and E270 are the catalytic residues.*
 RL Biochemistry 36:15489-15500(1997).
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-GALACTOSIDIC
 CC LINKAGES IN APARTINGALACTANS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 53 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: X91885; CAA62990.1; -
 FW Hydrolyase; Glycosidase; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 376
 FT APTINGALACTAN ENDO-1,4-BETA-
 FT GALACTOSIDASE.
 FT PROTON DONOR.
 FT ACID SITE 161 161
 FT ACID SITE 270 270
 FT MUTAGEN 161 161
 FT MUTAGEN 270 270
 FT SEQUENCE 376 AA; 42315 MW; BECAFE7243C6CAE10 CRC64;

Query Match 17.3%; Score 98; DR 1; Length 376;

Best Local Similarity 43.5%; Pred. No. 4,290-03;
 Matches 10, Conservative 6, Mismatches 6, Indels 1, Gaps 1;

DB 344 WENAFEDATKNNKNDLPAFLFFK 366
 9 W-TIOPFANAKQSTSLPPIPIPK 30

RESULT 3
 ID NDF2_HUMAN STANDARD: PFI: 382 AA.
 AC Q15784;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE NEUROGENIC DIFFERENTIATION FACTOR 2.
 GN NEUROD2.
 OS Homo sapiens (human).
 CC Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Mollusca;
 CC Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
 RN (1)
 RC SEQUENCE FROM N.A.
 RX MEDLINE: 96413331.
 RA McCormick M.B., Tamimi E.M., Snider L., Asakura A., Bergstrom D.,
 Tappero S.J.;
 *Neurod2 and neurod3: distinct expression patterns and
 RT transcriptional activation potentials within the neurod gene
 RT family.*
 RL Mol. Cell. Biol. 16:5747-5800(1996).
 RN (2)
 RV DIVISIONS TO 336-338.
 RA Tappero S.J., Tamimi E.M., McCormick M.B.,
 Smith C.D. (1994) In: The EMBL Outstation -
 CC FUNCTION: APPEARS TO MEDATE NEURONAL DIFFERENTIATION.
 CC -1- SIMILARITY: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
 CC BHLH PROTEIN.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "AVONAL" SUPERFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: U56481; AAC26058.1; -
 RX MEDLINE: 9601725; -
 RA PFM: PF00010; HLH: 1.
 DR PROSITE: PS00038; HELIX_LOOP_HELIX: 1.
 KW Transcription regulation; DNA-binding; Nuclear protein;
 KW Differentiation.
 FT DOMAIN 37 46
 FT DOMAIN 82 91
 FT DOMAIN 107 113
 FT DNA_BIND 122 133
 FT DOMAIN 134 174
 FT DOMAIN 282 285
 FT SEQUENCE 382 AA; 41361 MW; E33AA5C0B8371H47 CRC64;

Query Match 16.0%; Score 97; DR 2; Length 382;
 Best Local Similarity 29.7%; Pred. No. 4,050-02;
 Matches 11, Conservative 10, Mismatches 15, Indels 1, Gaps 1;

DB 215 FETEGACGACGPPGSGGPPAMHPYPPSPRIACAG 251
 33 DLAIETISTEFANNSPSYIPYIP-PSLSSSQI 68

RESULT 4
 ID NDF2_MOUSE STANDARD: PFI: 383 AA.

CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=OSTEOSARCOMA;
 RX MEDLINE: 91225006.
 RA Klefer M.C., Maslarsz E.F., Bauer D.M., Zaff T.
 ET "Isolation and molecular cloning of two new 16 kDa insulin-like
 RT growth factor binding proteins isolated from adult human serum."
 RI J. Biol. Chem. 266:9043-9049(1991).
 RN [2]
 RP SEQUENCE OF 2-240 FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE: 92049376.
 RA Shimasaki S., Gao L., Shimonaka M., Lina N.
 ET "Isolation and molecular cloning of insulin-like growth
 RT factor-binding protein-6."
 RI Mol. Endocrinol. 5:938-948(1991).
 RN [3]
 RP PARTIAL SEQUENCE OF 28-42.
 RX MEDLINE: 91207395.
 RA Andrews D.L., Rihman R.S.;
 ET "A novel human insulin-like growth factor binding protein secreted by
 RT osteoblast-like cells."
 RI Biochem. Biophys. Res. Commun. 176:213-218(1991).
 RN [4]
 RP FULL-LENGTH SEQUENCE OF 28-42
 RC TISSUE=CEREBROSPINAL FLUID;
 RX MEDLINE: 90005986.
 RA Roghani M., Hossainpour P., Isagaye P., Kalland A., Binaux M.;
 ET "Isolation from human cerebrospinal fluid of a new insulin-like
 RT growth factor-binding protein with a selective affinity for IGF II."
 RI FEBS Lett. 255:253-258(1989).
 RN [5]
 RP CARBOHYDRATE-BINDING SITES.
 RX MEDLINE: 98241382.
 RA Neumann G.M., Malinaro J.A., Bach L.A.;
 ET "Isolation of 2 glycosylation sites and partial characterization
 RT of carbohydrate structure and disulfide linkages of human
 RI insulin-like growth factor binding protein 6."
 RI Biochemistry 37:6572-6585(1998).
 RN [6]
 RP DISULFIDE BONDS.
 RX MEDLINE: 99279603.
 RA Neumann G.M., Bach L.A.;
 ET "The N-terminal disulfide linkages of human insulin-like growth
 RT factor-binding protein-6 (hIGFBP-6) and hIGFBP-1 are different as
 RI determined by mass spectrometry."
 RI J. Biol. Chem. 274:14587-14594(1999).
 RN [7]
 RP FUNCTION: IGF-BINDING PROTEINS PROMOTE THE HALF-LIFE OF THE IGFs
 ET AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
 RT PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE
 RI INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- PTM: O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
 CC PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL consortium -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, changed by and for commercial
 CC entities (except for a license agreement (See <http://www.ebi.ac.uk/infocentre/copyright.htm>)
 CC or send an email to license@ebi.ac.uk)
 CC
 DR EMBL: M62402; AAR06187.1;
 DR EMBL: M69054; AAB88070.1;
 DR PIR: PH0144; PH0144.
 DR PIR: S05604; S05604.
 DR PIR: A39842; A39842.
 DR MIM: 146735;
 DR PFAM: PF00219; IGF3P.1.
 DP

RP PFAM: PF00086; thyroglobulin_1; 1.
 DR PROSITE: PS00222; IGF_BINDING, FALSE_NEG.
 DR PROSITE: PS00484; THYROGLOBULIN_1; 1.
 KW Growth factor binding; Signal; Glycoprotein;
 FT SIGNAL 1 24
 FT SIGNAL 25 240
 ET DOMAIN 186 234
 ET DISULFID 29 32
 ET DISULFID 40 44
 ET DISULFID 57 63
 ET DISULFID 71 84
 ET DISULFID 78 104
 ET DISULFID 163 190
 ET DISULFID 201 212
 ET DISULFID 214 234
 ET CARBOHYD 126 126
 ET CARBOHYD 144 144
 ET CARBOHYD 145 145
 ET CARBOHYD 146 146
 ET CARBOHYD 152 152
 ET CONFLICT 2 2
 SQ SEQUENCE 240 AA; 25422 MW; 2A70827F0A9C99 CRC64;
 Query Match 14.4%; Score 82; DB 1; Length 240;
 Exact local similarity 50.0%; Ident. No. 1,6-e+00;
 Matches 11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
 ID 7 UPPHUAUUUASPGGALAK 48
 C 23 LEFLLFLNLLIAVTEETIER 44
 RESULT 12
 ID IGH4_HSV1 STANDARD; PRT; 1298 AA.
 AC P08392;
 DI 01 AUG-1988 (Ref. 08, Created)
 ET 01 AUG-1988 (Ref. 08, Last sequence update)
 DT 15 JUN-1998 (Ref. 36, Last annotation update)
 DE SEARS ACING: IPANSCRIPTONAL PROTEIN IGH4 (TRANSCRIPTIONAL ACTIVATOR
 DE IE175) (ALPHA-4 PROTEIN).
 GN IGH4 OR IE175 OR RSL.
 OS Herpes simplex virus (type 1; strain 17).
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Alphaherpesvirinae; Simplexvirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 88274327.
 RA McGeech P.T., Enirympo N.A., Parison A.T., Payne A., Frame M.C.,
 ET McNab D., Perry L.J., Scott J.E., Taylor P.;
 RT "The complete DNA sequence of the long unique region in the genome of
 RI herpes simplex virus type 1."
 RI J. Gen. Virol. 69:1531-1574(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 85148504.
 RA McGeech D.J., Dolan A., Donald S., Brauer D.H.K.;
 ET "Complete DNA sequence of the short repeat region in the genome of
 RI herpes simplex virus type 1."
 RI Nucleic Acids Res. 14:1727-1745(1986).
 RN [3]
 RP DNA-BINDING DOMAIN.
 RX MEDLINE: 90171974.
 RA Wu C.-T., Wittek K.W.;
 ET "Codons 262 to 490 from the herpes simplex virus ICP4 gene are
 RT sufficient to encode a sequence-specific DNA binding protein."
 RI Nucleic Acids Res. 18:531-538(1990).
 RN [4]
 RP INFLUENCE OF PHOSPHORYLATION ON FUNCTION.
 RX MEDLINE: 91122847.
 RA Pavasavillan A.G., Wilcox K.W., Silverstein S.J.;
 ET "The interaction of ICP4 with cell-injected cell factors and its
 RT state of phosphorylation modulate differential recognition of leader


```

PN 121
RP SEQUENCE FROM N.A.
KA STRAIN-VF5:
KA DECKERT G., WARREN P.V., GAASERLAND T., YOUNG W.G., LENIX A.L.,
KA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., ADJAY M., HUBER R.,
KA FELDMAN R.A., SHOKI J.M., OLSON G.J., SWANSON R.V.:
KL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DK EMBL: AB000715; AAC07049.1;
DK PRAM: PF00035; dsmc 1;
DK PFAM: PF00636; Ribonuclease 1;
SQ SEQUENCE 221 AA: 26100 MW: 54921901 pI=7.12.

Query Match 15.88: Score 87; DB 2; Length 221;
Host Local Similarity 24.58; Prod. No. 2,890 01;
Matches 12; Conservative 17; Mismatches 18; Indels 2; Gaps 2;

TH 105 GIVEFAMAAVYIDSRFQANTFETLYFLFEPTLSAIREQPEKVE 154
QY 1 MGVPSTLWTFQFANARQST-SLPPL-LFKMLLAETSLERAHN 17

RESULT 14
ID 067319 PRELIMINARY: PRT: 219 AA.
AC 067319:
I1 01-AUG-1998 (IEMBLrel. 07, created)
I1 01-AUG-1998 (IEMBLrel. 07, last sequence update)
I1 01-NOV-1998 (IEMBLrel. 08, last annotation update)
IE HYPOTHETICAL 25.0 KD PROTEIN.
GN AO_1287.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
PN 111
RP SEQUENCE FROM N.A.
KA STRAIN-VF5:
KA MEDLINE: 98156566.
KA DECKERT G., WARREN P.V., GAASERLAND T., YOUNG W.G., LENIX A.L.,
KA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., ADJAY M., HUBER R.,
KA FELDMAN R.A., SHOKI J.M., OLSON G.J., SWANSON R.V.:
RT *The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.*
KL Nature 392:453-458(1998).
KN 121
RP SEQUENCE FROM N.A.
KA STRAIN-VF5:
KA DECKERT G., WARREN P.V., GAASERLAND T., YOUNG W.G., LENIX A.L.,
KA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., ADJAY M., HUBER R.,
KA FELDMAN R.A., SHOKI J.M., OLSON G.J., SWANSON R.V.:
KL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DK EMBL: AB000712; AAC07280.1;
DK Hypothetical protein.
SQ SEQUENCE 219 AA: 25050 MW: 0439273 CIRC12;

Query Match 15.1K: Score 86; DB 2; Length 219;
Host Local Similarity 43.88; Prod. No. 5,560 01;
Matches 14; Conservative 6; Mismatches 19; Indels 2; Gaps 2;

ID 13 VESLFLFLLSLAT-FPKLLFDRLI 43
QY 4 VESL-LWTFQFANARQSTSLPPLFKMLL 34

RESULT 15
ID 024175 PRELIMINARY: PRT: 389 AA.
AC 024175:
I1 01-JAN-1998 (IEMBLrel. 05, created)
I1 01-JAN-1998 (IEMBLrel. 05, last sequence update)
I1 01-NOV-1999 (IEMBLrel. 12, last annotation update)
GN RFL.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Oryza.

```

```

KN 111
RP SEQUENCE FROM N.A.
KA STRAIN CV: TORIDE; 1188DE LEVEL-PRING PANICLE;
KA KYOZAKA J., KONISHI S., NEM T. F., IZAWA T., SUMIYAMA T.:
KL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DK EMBL: AB005620; AA21547.1;
DK MIMDID: 25485; 075411479; 21.48
SQ SEQUENCE 389 AA: 42500 MW: 00509860 CIRC12;

Query Match 14.8K: Score 84; DB 10; Length 389;
Host Local Similarity 41.79; Prod. No. 1,120 00;
Matches 13; Conservative 4; Mismatches 19;

DB 225 EVTELETVAKANN...
QY 34 LAETSLERAHNSYTFHHVPS-L-ISSQCLERH 14

Search completed: Fri Aug 25 17:00:52 2000
Job time : 83 secs.

```



CC The sequence represents a tomato ripening-associated CM-cellulase
CC (endo-1,4-beta-D-glucanase, EC-3.2.1.4), encoded by the tcl3 gene.
CC The encoding sequence has been isolated from a ripe tomato fruit
CC cDNA library in Escherichia coli using probes based on highly
CC conserved regions from bear and avocado glucanases. The cellulase
CC sequence may be used in construction of antisense DNA expression
CC cassettes, which may be inserted into vector constructs for
CC Agrobacterium-mediated gene transfer to e.g. tomato plants. The
CC antisense construct controls the activity of the gene during fruit
CC ripening, and reduces fruit softening by inhibiting glucanase
CC activity.
SU Sequence 452 AA:

Query Match 20.68; Score 69; DB 1; Length 452;
Best local Similarity 36.48; Pred. No. 5,95e+01;
Matches 8; Conservative 7; Mismatches 6; Indels 1; Gaps 1;

Db 115 SLVFNKNKAYSQ-KLVHGAFTL 135
UY 13 SLTFNEASTVCGKFTINISVL 34

Search completed: Fri Aug 25 17:44:08 2000
Job time : 23 secs.

 M E S S E N G E R

Release 3.1A John F. Collins, BioComputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Meshc.jp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Aug 25 17:47:13 2000; MasPar time 2.47 seconds
 Tabular output not generated. 279,189 Million cell updates/sec

Title: US-09-257-585-2
 Description: (1-45) from US09257585.pep
 Perfect Score: 335
 Sequence: 1 MINKEYVSSTGSH1HNASVQCKHINQSVLPPDQNSVSKR 45

Scoring Table:
 PAM 150
 Gap 11

Searched: 152433 seqs, 15429240 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: a-15500MB 2:56:00MB 3:4:00MB 4:PT_COMB 5:backlist

Statistics: Mean 22.710; Variance 85.911; scale 0.265

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	73	21.8	484	US-09-276-	Sequence 9, Applicatio	1.86e+01
2	73	21.8	484	US-08-252-	Sequence 9, Applicatio	1.86e+01
3	72	21.5	463	US-08-162-	Sequence 9, Applicatio	2.27e+01
4	72	21.5	740	US-08-696-	Sequence 2, Applicatio	2.27e+01
5	71	21.5	670	US-08-307-	Sequence 2, Applicatio	2.27e+01
6	70	20.9	119	US-08-340-	Sequence 15, Applicatio	3.37e+01
7	70	20.9	120	US-08-274-	Sequence 15, Applicatio	3.37e+01
8	70	20.9	359	US-08-865-	Sequence 4, Applicatio	3.37e+01
9	70	20.9	359	US-08-137-	Sequence 4, Applicatio	3.37e+01
10	70	20.9	610	US-08-720-2	Sequence 4, Applicatio	3.37e+01
11	70	20.9	610	US-08-365-	Sequence 3, Applicatio	3.37e+01
12	69	20.6	452	US-08-271-	Sequence 6, Applicatio	4.11e+02
13	69	20.6	452	US-08-444-	Sequence 6, Applicatio	4.11e+02
14	66	19.7	120	US-08-274-	Sequence 37, Applicatio	7.38e+01
15	65	19.4	128	US-08-928	Sequence 3, Applicatio	8.95e+01
16	65	19.4	490	US-08-630-	Sequence 6, Applicatio	8.95e+01
17	64	19.1	906	US-08-734-	Sequence 48, Applicatio	1.08e+02
18	63	18.8	184	US-07-828-	Sequence 48, Applicatio	1.08e+02
19	62	18.5	34	US-08-603-	Sequence 42, Applicatio	1.59e+02
20	62	18.5	34	US-07-603-	Sequence 11, Applicatio	1.59e+02
21	62	18.5	417	US-08-277	Sequence 8, Applicatio	1.59e+02
22	62	18.5	417	US-08-410	Sequence 8, Applicatio	1.59e+02
23	62	18.5	417	US-07-649-	Sequence 8, Applicatio	1.59e+02

24	62	18.5	434	US-07-762-	Sequence 1
25	62	18.5	556	US-08-407-	Sequence 2
26	62	18.5	2516	US-08-474	Sequence 2
27	61	18.2	128	US-08-928-	Sequence 1
28	61	18.2	35	US-08-433-	Sequence 4
29	61	18.2	36	US-08-989-	Sequence 4
30	61	18.2	593	US-08-111-	Sequence 4
31	61	18.2	619	US-08-818	Sequence 1
32	61	18.2	931	US-08-463	Sequence 7
33	61	18.2	1129	US-08-628	Sequence 1
34	61	18.2	1176	US-08-445-	Sequence 1
35	61	18.2	1176	US-08-456	Sequence 2
36	61	18.2	1176	US-08-456	Sequence 2
37	61	18.2	1176	US-08-456	Sequence 2
38	61	18.2	1176	US-08-456	Sequence 2
39	61	18.2	1176	US-08-456	Sequence 2
40	61	18.2	1176	US-08-456	Sequence 2
41	61	18.2	1176	US-08-456	Sequence 2
42	61	18.2	1176	US-08-456	Sequence 2
43	61	18.2	1176	US-08-456	Sequence 2
44	61	18.2	1176	US-08-456	Sequence 2
45	61	18.2	2160	US-08-514-	Sequence 2

ALIGNMENTS

RESULT 1
 US-09-276-197-9 SEQUENCE 9, Application 08/02/97 197
 484 AA

Sequence 9, Application 08/02/97 197

Patent No. 6040428

GENERAL INFORMATION:

APPLICANT: ROBERT ROSS, INC.

APPLICANT: EVANOS, MARK J.

APPLICANT: MATTHEW A.

TITLE OF INVENTION: PROTEIN E-SELECTIN

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: 25 Solway Park, Box 15

CITY: New Bedford

STATE: Connecticut

COUNTRY: USA

ZIP: 06511

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 720 Kb storage

COMPUTER: PC compatible

SOFTWARE SYSTEM: DOS 5.0

SOFTWARE: MEGALITH 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: 09/02756,197

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

AFFILIATION NUMBER: 09/02756,197

FILING DATE:

AUTHOR/ACCESSION NUMBER:

NAME: Fiedel, Scott A.

REGISTRATION NUMBER: 48,449

REFERENCE/DOCKET NUMBER: ALX-148

TELECOMMUNICATION INFORMATION:

TELEPHONE: (203) 772-1791

TELEFAX: (203) 772-0655

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 484 amino acids

00 TYPE: amino acid
00 STRANDEDNESS: Single
00 TOPOLOGY: Linear
00 MOLECULE TYPE: cDNA to mRNA
00 DESCRIPTION: Predicted amino acid sequence of
00 SEQUENCE: 484 AA: 52567 MW: 1238841 CN:

Query Match 21.8% Score 73; DB 3; Length 484;
Host Local Similarity 28.6%; Pred. No. 1,866+01;
Matches 6; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Db 31 TMTPDASAVCOQRYTHLVAI 51
13 STTFNEASTVCGKFINGTSV 33

RESULT 2 STANDARD: PRT: 484 AA.

XX US-08-252-4930-9
XX xxxxxx

DE Sequence 9, Application US/0812524930

00 Sequence 9, Application US/082524930

00 Patent No. 5891645

00 GENERAL INFORMATION:

00 APPLICANT: Rollins, Scott

00 APPLICANT: Rollins, Russell P.

00 APPLICANT: Evans, Mark J.

00 APPLICANT: Mattis, Louis A.

00 TITLE OF INVENTION: PORCINE E-SHEDDIN

00 NUMBER OF SEQUENCES: 9

00 CORRESPONDENCE ADDRESS:

00 ADDRESSEE: Seth A. Fidel

00 STREET: 25 Seaside Park, Box 15

00 CITY: New Haven

00 STATE: Connecticut

00 COUNTRY: USA

00 ZIP: 06511

00 COMPUTER READABLE FORM:

00 MEDIUM TYPE: 3.5 inch, 750 KB storage

00 COMPUTER: PC compatible

00 OPERATING SYSTEM: IFS 6.2

00 SOFTWARE: Wordperfect 6.0

00 CURRENT APPLICATION DATA:

00 APPLICATION NUMBER: US/082524930

00 FILING DATE: June 1, 1994

00 CLASSIFICATION: 435

00 PRIOR APPLICATION DATA:

00 APPLICATION NUMBER:

00 FILING DATE:

00 ATTORNEY/AGENT INFORMATION:

00 NAME: Fidel, Seth A.

00 REGISTRATION NUMBER: 38,449

00 REFERENCE/BOOK NUMBER: ALX-138

00 TELECOMMUNICATION INFORMATION:

00 TELEPHONE: (203) 772-1790

00 TELEFAX: (203) 772-1655

00 INFORMATION FOR SEQ ID NO: 9:

00 SEQUENCE CHARACTERISTICS:

00 LENGTH: 484 amino acids

00 TYPE: amino acid

00 STRANDEDNESS: Single

00 TOPOLOGY: Linear

00 MOLECULE TYPE: cDNA to mRNA

00 DESCRIPTION: Predicted amino acid sequence of

00 SEQUENCE: 484 AA: 52567 MW: 1238841 CN:

Query Match 21.8% Score 73; DB 2; Length 484;

Best Local Similarity 28.6%; Pred. No. 1,866+01;
Matches 6; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Db 31 TMTPDASAVCOQRYTHLVAI 51
13 STTFNEASTVCGKFINGTSV 33

RESULT 4 STANDARD: PRT: 483 AA.

XX US-08-162-4026-9

XX xxxxxx

DE Sequence 9, Application US/081624026

00 Sequence 9, Application US/081624026

00 Patent No. 5972337

00 GENERAL INFORMATION:

00 APPLICANT: CERIANI, ROBERTO L.

00 APPLICANT: PETERSON, JERRY A.

00 APPLICANT: LAROCKA, DAVID J.

00 TITLE OF INVENTION: 46 KILTON HUMAN MILK FAT

00 NUMBER OF SEQUENCES: 29

00 CORRESPONDENCE ADDRESS:

00 ADDRESSEE: Pretty, Schroeder & Poplawski

00 STREET: 444 South Flower St., 19th floor

00 CITY: Los Angeles

00 STATE: CA

00 COUNTRY: USA

00 ZIP: 90071

00 COMPUTER READABLE FORM:

00 MEDIUM TYPE: Diskette

00 COMPUTER: IBM compatible

00 OPERATING SYSTEM: DOS

00 SOFTWARE: FASTSEQ for Windows Version 2.0

00 CURRENT APPLICATION DATA:

00 APPLICATION NUMBER: US/081624026

00 FILING DATE: 03-DEC-1993

00 CLASSIFICATION: 435

00 PRIOR APPLICATION DATA:

00 APPLICATION NUMBER:

00 FILING DATE:

00 ATTORNEY/AGENT INFORMATION:

00 NAME: Amzol, Viviana

00 REGISTRATION NUMBER: 30,940

00 REFERENCE/BOOK NUMBER: P. 36-15

00 TELECOMMUNICATION INFORMATION:

00 TELEPHONE: 213-622-9700

00 TELEFAX: 213-489-4210

00 INFORMATION FOR SEQ ID NO: 9:

00 SEQUENCE CHARACTERISTICS:

00 LENGTH: 463 amino acids

00 TYPE: amino acid

00 STRANDEDNESS: unknown

00 TOPOLOGY: unknown

00 MOLECULE TYPE: Peptide

00 SEQUENCE: 463 AA: 51416 MW: 1096958 CN:

Query Match 21.5% Score 72; DB 2; Length 463;
Host Local Similarity 29.4%; Pred. No. 2,270+01;
Matches 10; Conservative 10; Mismatches 13; Indels 1; Gaps 1;

Db 16 CASGLFAASDECDSSLTCTGTCGQNDIYC 49
11 COSITFNASTVCGKFINGTSV 43

RESULT 4 STANDARD: PRT: 730 AA.


```

07      ZITE: 10920
08      COMPUTER READABLE FORM:
09          MEDIUM TYPE: Floppy disk
10      COMPUTER: IBM PC compatible
11      OPERATING SYSTEM: pc-dos/ms-dos
12      SOFTWARE: Patent In Release #1.0, Version #1.30
13      CURRENT APPLICATION DATA:
14      APPLICATION NUMBER: US/08/340,639A
15      FILING DATE: 16 NOV 1994
16      CLASSIFICATION: 51A
17      PRIORITY INFORMATION:
18      APPLICATION NUMBER: US 08/008,459
19      FILING DATE: 25 JAN 1993
20      ATTORNEY/AGENT INFORMATION:
21      NAME: Gannison, Laro
22      REGISTRATION NUMBER: 38,479
23      REFERENCE TO OTHER APPLICAT.: 04-04
24      TELECOMMUNICATIONS INFORMATION:
25      TELEPHONE: 212 596 9030
26      TELEXFAX: 212-596-9090
27      INFORMATION FOR SEQ ID NO: 15:
28      SEQUENCE CHARACTERISTICS:
29      LENGTH: 115 amino acids
30      TYPE: amino acid
31      STRANDEDNESS:
32      TOPOLOGY: linear
33      MOLECULE TYPE: protein
34     SEQUENCE: 119 AA; 14118 MW; 73163 CN;
35
36 Query Match: 20.9%; Score 70; DB 1; Length 11%;
37 Best Local Similarity 28.6%; Ered.No. 3.47e+01;
38 Matches: 6; Conservative: 9; Mismatches: 6; Indels: 0; Gaps: 0;
39
40 DB: 9 AMIDEASAYCOOKYHLEVAL 29
41 : ::::| | | | | : : :
42 QY: 13 SLTFNEASTVCQCKFNITNSV 33
43
44 RESULT: 7
45 ID: US-08-274-661R-36 STANDARD; PRT: 120 AA.
46 XX: xxxxxx
47 XX:
48 DE: Sequence 36, Application US/08274661R
49 XX:
50      Sequence 36, Application US/08274661B
51      Patent No. 5593882
52      GENERAL INFORMATION:
53      APPLICANT: Erbo, David V.
54      APPLICANT: Laskey, Laurence A.
55      TITLE OF INVENTION: Selection Variants
56      NUMBER OF SEQUENCES: 49
57      CORRESPONDENCE ADDRESS:
58      ADDRESSEE: Genentech, Inc.,
59      STREET: 460 Point San Bruno Blvd
60      CITY: South San Francisco
61      STATE: California
62      COUNTRY: USA
63      ZIP: 94080
64      COMPUTER READABLE FORM:
65      MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
66      COMPUTER: IBM PC compatible
67      OPERATING SYSTEM: PC-DOS/MS DOS
68      SOFTWARE: Winpatin (Genentech)
69      CURRENT APPLICATION DATA:
70      APPLICATION NUMBER: US/08/274,661R
71      FILING DATE: 13-Jul-1994
72      CLASSIFICATION: 42A
73      PRIOR APPLICATION DATA:
74      APPLICATION NUMBER: 07/956701

```

```

01 FILING DATE: 10/31/1992
02 ATTORNEY/AGENT INFORMATION:
03 NAME: Dreger, Ginger R.
04 REGISTRATION NUMBER: 43,045
05 REFERENCE/DOCKET NUMBER: 761101
06 TELECOMMUNICATION INFORMATION:
07 TELEPHONE: 415/225-3216
08 TELEFAX: 415/952-9881
09 TELE: 910/371-7168
10 INFORMATION FOR SEQ ID NO: 36:
11 SEQUENCE CHARACTERISTICS:
12 LENGTH: 120 amino acids
13 TYPE: Amino Acid
14 TOPOLOGY: Linear
15 SEQUENCE: 120 AA; 14189 MW; 72284 CNI;
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```





```

ACCESSIONS      02-Jul-1998
REFERENCE       D70410
AUTHORS         Decker, G.; Warren, P.V.; Maasterland, T.; Young, M.G.;
                  Lenox, A.L.; Graham, D.E.; Overbeck, R.; Sheph, M.A.;
                  Keller, M.; Augay, M.; Huber, R.; Feldman, K.A.; Short,
                  J.M.; Olson, G.J.; Skason, R.V.
#journal        Nature (1998) 392:353-358
#title          The complete genome of the hyperthermophilic bacterium
                  Aquifex aeolicus.
#cross-references MIMD:98196666
#accession      D70410
#status         preliminary; nucleic acid sequence not shown;
                  translation not shown
#molecule_type DNA
#residues       1458 #label AQP
#cross-references AF000742; M17-298704; P1D-298706; GB-AE000657
#experimental_source strain: VFS
GENETICS
#gene           catA
SUMMARY          #length 458 #molecular-weight 52508 #checksum 3184
Query Match     21.5% Score 72; EB 2; Length 458;
Best Local Similarity 39.2%; Pred. No. 2,146+00;
Matches 13; Conservative 11; Mismatches 15; Indels 4; Gaps 4;
ID 364 MKKWESESY-EVI-GETPHVQKGFVSKREFTLITLTKAL 404
ID 1 MTNKVFKSTGSLTFENASYVQGR-FLNGTS-VLPLDLSV 41

```

```

ORGANISM        #formal_name Arabidopsis thaliana #nom
DATE            16-Jul-1999 #sequence_revision 14
ACCESSIONS      T10575
REFERENCE       216533
AUTHORS         Heyan, M.; Lehmuth, N.; Quail, M.; Hart,
                  K.F.X.; Lamke, K.; Scheller, G.
#submission     Submitted to the Protein Sequence Data
#accession      T10575
#status         preliminary
#molecule_type DNA
#residues       1456 #label HEV
#cross-references EMBL:AL013992; USEDB:GNDL 42
#experimental_source cultivar Columbia; BAC clone
GENETICS        AFS:PF294.17
#gene           map_position 4
SUMMARY          #length 608 #molecular-weight 7617
Query Match     21.5% Score 72; EB 2; Length
Best Local Similarity 27.7%; Pred. No. 2,146+00;
Matches 8; Conservative 10; Mismatches 10;
ID 20 MKNVITRAVTCGRPR-LSQVYVTS-SPH 49
ID 1 XTPVYVPEPQVLT-TRCAVYV-20 EPI 28

```

Search completed: Fri Aug 25 17:14:14 2000
Job time : 17 secs.


```

CC EMBL: X07805; CAA0663.2;
CC PIR: G30045; VCLJ04;
CC HIV: X07805; EVISA01V;
CC PRAM: PF00516; GP120; 1;
CC PIR: PF00517; GP41; 1;
CC ATUS: coat protein; Polyprotein; Glycoprotein; Transmembrane;
CC SIGNAL.
CC
CC ID: SIGNAL. 1 218
CC F1 CHAIN 219 536 EXTERIOR MEMBRANE GLYCOPROTEIN,
CC F1 SITE 537 865 TRANSMEMBRANE GLYCOPROTEIN,
CC F1 SITE 770 770 IN-FRAME TERMINATION CODON.
CC F1 CARBOHYD 45 35 POTENTIAL.
CC F1 CARBOHYD 68 68 POTENTIAL.
CC F1 CARBOHYD 117 117 POTENTIAL.
CC F1 CARBOHYD 150 150 POTENTIAL.
CC F1 CARBOHYD 165 165 POTENTIAL.
CC F1 CARBOHYD 195 195 POTENTIAL.
CC F1 CARBOHYD 198 198 POTENTIAL.
CC F1 CARBOHYD 210 210 POTENTIAL.
CC F1 CARBOHYD 252 252 POTENTIAL.
CC F1 CARBOHYD 255 255 POTENTIAL.
CC F1 CARBOHYD 266 266 POTENTIAL.
CC F1 CARBOHYD 276 276 POTENTIAL.
CC F1 CARBOHYD 282 282 POTENTIAL.
CC F1 CARBOHYD 294 294 POTENTIAL.
CC F1 CARBOHYD 305 305 POTENTIAL.
CC F1 CARBOHYD 316 316 POTENTIAL.
CC F1 CARBOHYD 373 373 POTENTIAL.
CC F1 CARBOHYD 414 414 POTENTIAL.
CC F1 CARBOHYD 451 451 POTENTIAL.
CC F1 CARBOHYD 488 488 POTENTIAL.
CC F1 CARBOHYD 491 491 POTENTIAL.
CC F1 CARBOHYD 645 645 POTENTIAL.
CC F1 CARBOHYD 661 661 POTENTIAL.
CC
CC SEQID: 865 AA; 39024 MM; bcdh090010b095 chr054;
CC
CC Query Match 22.1% Score 74; DB 1; Length 865;
CC Post Local Similarity 44.0%; Proc. No. 2,910-01;
CC Matches 11; Conservative 4; Mismatches 9; Indels 1; Gaps 1;
CC
CC ID: 178 KKNYSVWMDAETKKNK-TNLS 201
CC Y 8 KSTVSSTFENASTVWYKPTNLS 22
CC
CC RESULT 9
CC ID: YIP4 YEAST STANDARD PRT: 493 AA.
CC AC: PA0454;
CC DT: 01-FEB-1995 (Rel. 4); Created;
CC DT: 01-FEB-1995 (Rel. 4); Last sequence update;
CC DT: 01-FEB-1995 (Rel. 4); Last annotation update;
CC DE: HYDROPHETICAL 45.1 KDa protein in 100% DNA3 INTERCEPT REGION.
CC GN: YIL154W.
CC OS: Saccharomyces cerevisiae (Baker's yeast).
CC OC: Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
CC NC: Saccharomycetaceae; Saccharomycos.
CC
CC SEQUENCE FROM N.A.
CC AC: STRAIN-S288C / AB972;
CC AC: Barrett R.G., Padgett K., Bankier A.T., Bowman S., Brown D.,
CC Gutherie C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
CC Gough S., Hamlyn N., Horsnell T.S., Hunt S., Jangels K., Jones M.,
CC Louis E., Lye G., Monte S., Monte T., Odell C., Pearson D.,
CC Rajandram M.A., Riles L., Rowley N., Skelton J., Smith V.,
CC Walsh S.V., Whitehead S.;
CC Submitted (DEC-1994) to the EMBL/GenBank/DBP databases
CC
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entries requires a license agreement (See http://
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: Z38059; CAA6125.2;
CC PIR: S48381; S48381;
CC Hypothetical protein.
CC
CC SEQUENCE 493 AA; 493 AA; 215284559/990 b.0;
CC
CC Query Match 21.9% Score 73; DB 1; Length 493;
CC Post Local Similarity 44.0%; Proc. No. 4,400-01;
CC Matches 8; Conservative 11; Mismatches 4;
CC
CC ID: 242 YKNTNYSGQINTNPK MATCH 65
CC Y 16 YKNTNYSGQKTHNDSVPTN 49
CC
CC RESULT 10
CC ID: LEM2 PIG STANDARD PRT: 484 AA.
CC AC: P98110;
CC DT: 01-FEB-1996 (Rel. 4); Created;
CC DT: 15-JUL-1998 (Rel. 4); Last sequence update;
CC DE: E SELECTIN PRECURSOR (ENDOTHELIAL LEUKOCYTE ADHESION MOLECULE 1) (CD62E).
CC DE: (CD62E).
CC OS: Sus scrofa (Pig).
CC OC: Eukaryota; Metazoa; Chordata; Artiodactyla; Artiodactyla;
CC NC: Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus;
CC GN: LEM2;
CC AC: MEDLINE: 95071392;
CC AC: Rollins S.A., Evans M.L., Johnson K.K., Elliot E.J.,
CC Malis L.A., Rolher R.P.;
CC RA: "Molecular and functional analysis of porcine E-selectin: a potential role in xenograft rejection."
CC RT: Biochem. Biophys. Res. Commun. 204:763-771(1994).
CC
CC ID: 121
CC AC: SEQUENCE FROM N.A.
CC AC: TISSUE-AORTIC ENDOTHELIAL;
CC AC: MEDLINE: 94271236;
CC AC: Takeda Y.T.M., Baskind J., Pederson M.K.;
CC RA: "Cloning and expression kinetics of porcine E-selectin molecule."
CC RT: Biochem. Biophys. Res. Commun. 201:805-809(1994).
CC DE: "FUNCTION: EXPRESSED IN THE KIDNEY INDUCED BY ENDOTHELIAL MEDICINES THEIR BINDING TO APOMOYLES. THE LIGAND ELAM-1 IS STABLE LIGAND X (ALPHA(1-3)FUCOSYLTRANSFERASE) THAT ARE FOUND AT THE N-TERMINUS GLYCOPOLYDIPS).
CC DE: REJECTION AND BROADLY ASSOCIATED IN XENOGRAFT REJECTION. SIMILARITY TO OTHER SELECTINS/PIRAMS.
CC DE: SIMILARITY: CONTAINS 1 OF TYPE I/LECTIN FAMILY 1.
CC DE: SIMILARITY: CONTAINS 4 SUSHI (SUSHI) REPEATS: 1
CC DE: THE HUMAN SUSHI 1 AND 4 EQUIVALENTS.
CC
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entries requires a license agreement (See http://
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: L39076; AAA6154.1;
CC PIR: U08350; AAA2154.1;
CC HSP: P16581; PSL;
CC PRAM: PF00008; EGF; 1;

```

DR PFAM: PF00059; lectin; 1.
 DR PFAM: PF00084; sushi; 4.
 DR PRINTS: PR00343; SELECTIN.
 DR PROSITE: PS00022; EGF_1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS09415; C-TYPE LECTIN_1; 1.
 DR PROSITE: PS09041; C-TYPE LECTIN_2; 1.
 DR Cell adhesion: Transmembrane; Glycoprotein; EGF-like domain; lectin;
 KM Selectin: Signal: Sushi: Repeat
 ET SIGNAL 1 22
 ET CHAIN 23 484
 ET TRANSMEM 24 429
 ET DOMAIN 430 451
 ET DOMAIN 452 484
 ET DOMAIN 459 484
 ET DOMAIN 140 176
 ET DOMAIN 180 421
 ET REPEAT 180 236
 ET REPEAT 239 299
 ET REPEAT 302 362
 ET REPEAT 365 421
 ET REPEAT 41 139
 ET DISULFID 112 131
 ET DISULFID 144 155
 ET DISULFID 149 164
 ET DISULFID 166 175
 ET DISULFID 181 222
 ET DISULFID 208 235
 ET DISULFID 240 285
 ET DISULFID 271 298
 ET DISULFID 303 348
 ET DISULFID 334 361
 ET DISULFID 366 407
 ET DISULFID 394 420
 ET CARBOHYD 61 61
 ET CARBOHYD 65 65
 ET CARBOHYD 79 79
 ET CARBOHYD 160 160
 ET CARBOHYD 201 201
 ET CARBOHYD 254 254
 ET CARBOHYD 376 376
 ET CARBOHYD 400 400
 ET CONFLICT 253 253
 ET CONFLICT 313 313
 ET CONFLICT 322 321
 ET CONFLICT 327 327
 ET CONFLICT 363 363
 ET CONFLICT 384 384
 ET CONFLICT 461 484
 SEQUENCE 484 AA; 52567 MW; AFF7AFEE25C1FD13 Ck:04;
 Query Match 21.8%; Score 73; DB 1; Length 484;
 Host Local Similarity 28.6%; Pred. NO. 4,406 0;
 Matches 6; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
 DB 31 MITOCHONDRIAL
 QY 13 STIFFNESS/COGKFINITSV 33
 RESULT 11
 ID IF2 BORBO STANDARD; PRT: 882 AA.
 AC 051741.
 DI 15-DEC-1998 (Rel. 37, Created)
 DI 15-DEC-1998 (Rel. 37, last sequence update)
 DI 15-DEC-1998 (Rel. 37, last annotation update)
 DE TRANSLATION INITIATION FACTOR IF-2.
 GN INE2 OF PROBOL
 OS Borrelia burgdorferi (Lyme disease spirochete).
 NC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 RN 11
 SEQUENCE FROM N.A.

RC STRAIN-AVCC 35210 / B31;
 RA MEDLINE: 98065943.
 RA Fraser G.M., Gasjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lachiga P., White O., Ketchum K.A., Deason F., Hickey E.K., Gwin M.,
 RA Jacoby B., Tomb J.F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Fajlavay A.P., Eisenbach J., Stachert S., Hanson M.,
 RA van Vliet K., Palmer N., Adams M.D., Ocayao J.D., Weidman J.,
 RA Overlack T., Matthey L., McDonald M.D., Atchick P., Roman C.,
 RA Garland S., Fujii C., Cotton M.D., Host K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi* *;
 RI Nature 390:580-586(1997).
 CC - FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION
 CC OF PROTEIN SYNTHESIS IN VITRO, PROMOTES FORMYLETIONYL-TRNA FROM
 CC SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 90S
 CC RIBOSOMAL SUBUNIT. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP
 CC DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX (BY SIMILARITY).
 CC 1. SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC 2. SIMILARITY: BELONGS TO THE IF-2 FAMILY.
 CC -----
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see EBI's <http://www.ebi.ac.uk/seqdb/doc/seqdb.html>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: AB001179; AAC67154.1;
 DR ICB: HB0801;
 DR PFAM: PF00009; GTP_BIND; 1.
 DR PROSITE: PS01176; IF2; FALSE NEG.
 KM Initiation factor; Protein biosynthesis; GTP-binding.
 FT DOMAIN 383 531
 FT NP_BIND 389 396
 FT NP_BIND 435 439
 FT NP_BIND 489 492
 SEQUENCE 882 AA; 97795 MW; 0FA20D5B543041 CR664;
 Query Match 21.8%; Score 73; DB 1; Length 882;
 Host Local Similarity 25.0%; Pred. NO. 4,406 0;
 Matches 6; Conservative 11; Mismatches 7; Indels 0; Gaps 0;
 DB 563 KOVCSVTEEDLYVSGEVA 592
 QY 8 KSTCSLEFNEASTYCGKFINITSV 31
 RESULT 12
 ID MYM_MOUSE STANDARD; PRT: 463 AA.
 AC P21956; P97800;
 DI 01-AUG-1991 (Rel. 19, Created)
 DI 01-AUG-1991 (Rel. 19, last sequence update)
 DI 15-FEB-2000 (Rel. 39, last annotation update)
 DE LACTAMERIN PRECURSOR (MILK FAT GLOBULE-BOF FACTOR 8) (MGF18) (MFGM)
 DE (SPERM SURFACE PROTEIN SP47) (MP47).
 GN MYGB.
 OS Mus musculus (Mouse).
 NC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN 11
 SEQUENCE FROM N.A. AND SEQUENCE OF 23-35.
 RC TISSUE-MAMMARY GLAND;
 RX MEDLINE: 91046008.
 RA Stubbs J.D., Iekuts G., Singer K.L., Bui A., Yuzuki D.,
 RA Srinivasan U., Parry G.;
 RT Cloning of a mouse mammary epithelial cell surface protein
 RT reveals the existence of epidermal growth factor-like domains linked
 RT to factor VIII-like sequences.*;
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
 RN 12
 SEQUENCE OF 23-463 FROM N.A.

RA TISSUE-TESTIS;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN PROSPELPTID BINDING. ZOEA PELLUCIDA
 CC BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRAN PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL. SUBFACIES AND
 CC SPERMATOZOOM.
 CC -1- DEVELOPMENTAL STAGE: MRNA EXPRESSION IS DETECTABLE IN MAMMARY
 CC TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isr-sib.ch/announc/annoy>
 CC or send an email to license@sib.ch).
 CC
 DR EMBL: M88337; AAA9544.1;
 DR EMBL: Y11684; CAA2480.1;
 DR PIR: A36479; A36479.
 DR HSSP: P00740; 11XA.
 DR MG: MG1102768; M8388.
 DR PRAM: P000008; EGF_2.
 DR PRAM: P000754; F5_18_Type_C_2.
 DR PRINITE: P000010; EGFRLD.
 DR PROSITE: PS00022; EGF_1_2.
 DR PROSITE: PS01186; EGF_2_2.
 DR PROSITE: PS01285; F5/8C_1_2.
 DR PROSITE: PS01386; F5/8C_2_2.
 KM Signal: glycoprotein; Repeat: EGF-like domain; Milk.
 FT STONL 1 22
 FT CHAIN 23 461
 FT DOMAIN 24 41
 FT DOMAIN 64 108
 FT DOMAIN 148 303
 FT DOMAIN 308 464
 FT SITE 87 89
 FT DISULFID 28 49
 FT DISULFID 33 49
 FT DISULFID 51 60
 FT DISULFID 68 79
 FT DISULFID 73 96
 FT DISULFID 98 107
 FT DISULFID 148 403
 FT DISULFID 290 294
 FT DISULFID 308 463
 FT CARBOHYD 61 61
 FT CARBOHYD 266 266
 FT CARBOHYD 316 416
 FT CARBOHYD 426 426
 FT CONFLICT 40 40
 FT CONFLICT 35 45
 FT CONFLICT 110 147
 FT CONFLICT 168 168
 FT CONFLICT 196 196
 FT CONFLICT 309 309
 FT CONFLICT 395 395
 SQ SEQUENCE 463 AA: 51465 MW: D7B66CEFBBAV24D CRG64:
 Query Match 21.5% Score 72: FB 1: Length 463:
 Most Local Similarity 29.4% Fred. No. 6,61e-01:
 Matches 10: Conservative 10: Mismatches 14: Indels 1: Gaps 1:
 ID 16 CASSTFASSTPSSSTTSMAGTTSQNNRYC 49
 YV 11 GSTTFNEASTYQGRK INGVSLPPLDSVLC 43

RESULT 13
 ID S026_CABDI STANAN: 570 AA
 AC Q10579;
 DT 01-OCT-1996 (Rel. 14, created)
 DT 01-OCT-1996 (Rel. 14, last sequence update)
 DT 01-OCT-1996 (Rel. 14, last annotation update)
 DE SPERMATOCYTE PROTEIN SPE-26.
 GN SPE-26 OR R10H10.2.
 OS Caenorhabditis elegans.
 OC Eubacteria: Metazoa: Nematoda: Chromadorea: Rhabdi-
 OC Rhabditidae: peloderae: Caenorhabditis.
 RN [1]
 KC SEQUENCE FROM N.A. AND MUTAGENESIS.
 KC STRAIN-BRISTOL N2;
 RX MEDLINE: 95262900.
 RA Vakey J.F., Mulholland J., Mangan A.N., De P., Ma-
 RT "The Caenorhabditis elegans spe-26 gene is necessary
 RP spermatids and encodes a protein similar to the an-
 RP protein Kelch and encodes a protein similar to the an-
 RI genes dev. 9:1074-108 (1997)
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Kaidi J.S.
 KC Submitted (Apr. 1997) to the EMBL/GenBank/DBJ data-
 CC -1- FUNCTION: MAY BE INVOLVED IN THE SPERMATOCYTE
 CC POSSIBLY INTERACTING WITH ATRIN.
 CC -1- TISSUE SPECIFICITY: TESTIS. IN BOTH SPERMATOCY-
 CC SPERMATOCYTES.
 CC -1- SIMILARITY: Homologous to mouse/murine family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced
 CC between the Swiss Institute of Bioinformatics and
 CC the European Bioinformatics Institute. There are
 CC use by non-profit institutions as long as its
 CC modified and this statement is not removed. Usage
 CC entities requires a license agreement (see <http://www.isr-sib.ch/announc/annoy>
 CC or send an email to license@sib.ch).
 CC
 DR EMBL: U20142; AAB0245.1;
 DR EMBL: Z70686; CAAV4610.1;
 DR WORMPEP: R10H10.2; C9606295.
 DR PRAM: P01344; Kelch.
 DR PRINITE: P00501; RHOCHITRAL.
 KM Cytoskeleton; Repeat:
 FT DOMAIN 244 462
 FT REPEAT 244 291
 FT REPEAT 292 341
 FT REPEAT 341 393
 FT REPEAT 394 443
 FT REPEAT 441 493
 FT MUTAGEN 360 360
 FT MUTAGEN 446 446
 FT MUTAGEN 446 446
 SQ SEQUENCE 570 AA: 4462 MW: 620928474.68467 26
 Query Match 21.5% Score 72: FB 1: Length
 Most Local Similarity 44.3% Fred. No. 6,61e-01:
 Matches 11: Conservative 9: Mismatches 10:
 ID 482 MRSQCAARVE-ENGVVQGR N 081AAVE 411
 YV 7 VKSTGSIIFNEASTY Q KINLVSLPPLD 38
 DE PROTEIN C4.

MESSAGE

(TW)

Release 4.1A John F. Collins, BioComputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_bp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:45:43 2000. Msrch time 7.94 seconds

Label output not generated. 392.845 Million. 11.14 updates/sec

Title: 308-09-257-585-2

Description: (1-45) from US09257585.pep

Perfect Score: 435

Sequence: 1 MNRKPVESLPSSTLHNASTVQKPEIMDSVLPKNSVRR 45

Scoring table: PAM 150

Gap 11

Searched: 2,5978 seqs, 694412 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Databases:

spiremb12
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mic 8:sp-orphanolo
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 42.965; Variance 47.601; scale 0.693

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	OR	ID	Description	Prod. No.
1	89	26	6	452	6	PTZB12.1 PEPTIT	1.00e+00
2	78	23	3	366	10	1 AMINOCYCLOPEPANE-7A	1.29e+01
3	78	23	3	401	5	PEP1.1 PEPTIDE	1.29e+01
4	78	23	3	421	5	01081.4 PEPTIDE	1.29e+01
5	78	23	3	664	11	11091.4 PEPTIDE	1.29e+01
6	76	22	7	1038	19	11091.4 PEPTIDE	1.29e+01
7	76	22	7	1038	19	CALCININ TRANSPORTING A	2.97e+01
8	76	22	7	2193	14	POLYPEPTIDE	2.97e+01
9	75	22	4	146	2	POLYPEPTIDE	2.97e+01
10	75	22	4	146	2	ELIA MAN	4.49e+01
11	75	22	4	146	2	RETINYL POLYACETATE	4.49e+01
12	73	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
13	73	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
14	73	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
15	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
16	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
17	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
18	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
19	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
20	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00

Result ID	Score	Match	Length	OR	ID	Description	Prod. No.
1	89	26	6	452	6	PTZB12.1 PEPTIT	1.00e+00
2	78	23	3	366	10	1 AMINOCYCLOPEPANE-7A	1.29e+01
3	78	23	3	401	5	PEP1.1 PEPTIDE	1.29e+01
4	78	23	3	421	5	01081.4 PEPTIDE	1.29e+01
5	78	23	3	664	11	11091.4 PEPTIDE	1.29e+01
6	76	22	7	1038	19	11091.4 PEPTIDE	1.29e+01
7	76	22	7	1038	19	CALCININ TRANSPORTING A	2.97e+01
8	76	22	7	2193	14	POLYPEPTIDE	2.97e+01
9	75	22	4	146	2	POLYPEPTIDE	2.97e+01
10	75	22	4	146	2	ELIA MAN	4.49e+01
11	75	22	4	146	2	RETINYL POLYACETATE	4.49e+01
12	73	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
13	73	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
14	73	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
15	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
16	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
17	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
18	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
19	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00
20	72	21	8	153	14	HYDROLYTIC RNA POLYMER	1.01e+00


```

25 Lactobacillus curvatus.
26 Bacteria: Firmicutes: Bacilli: Clostridia: Clostridiales:
27 Lactobacillales.
28 [1]
29 SEQUENCE FROM N.A.
30 VERKAI A., GOSALUBS M., PEREZ-MARTINEZ C.
31 Submitted (Jun 1999) to the EMBL/GenBank/DDBJ databases.
32 EMBL: U28164; AAC04153.1; -
33 SEQUENCE 146 AA; 16060 MW; A71DD94B CEC32;
34
35 Query Match
36 Best local similarity 22.4%; Score 75; DB 2; Length 146.
37 Matches 9; Conservative 8; Mismatches 10; Indels 1; Gaps 1;
38
39 DB 22 MIFENNAYAVPPLKGGEGT-POLQAKI 48
40 : : : : : : : : : : : : : : : :
41 14 IFNENAYVGGKFGINTSVLPIDNVCV 41
42
43 RESULT 10
44 ID 082218 PRELIMINARY: PRT: 466 AA.
45 AC 082218:
46 01-NOV-1998 (TEMBUREL_08, Created)
47 01-NOV-1998 (TEMBUREL_08, Last sequence update)
48 01-MAY-1999 (TEMBUREL_10, Last annotation update)
49
50 POTATIVE POLYGLUTAMINASE.
51 T29E15.10.
52 Arabidopsis thaliana (Mouse-ear cress).
53 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
54 euphyllophytes: Spermatophytes: Magnoliophyta: Angiosperms:
55 core eudicots: Rosidae: eucosids: Brassicales: Brassicaceae:
56 Arabidopsids.
57 [1]
58 SEQUENCE FROM N.A.
59 STRAIN CV-701061A:
60 ROUNSLEY S.D., LIN X., KAUH S., SHUA L.P., FUJI G.Y., MASON T.M.,
61 SHIN M., KONING C.M., FRASER C.M., SOMERVILLE G.R., WINTER J.D.,
62 Arabidopsis thaliana chromosome 11 BAC 129416 genomic sequence.
63 Submitted (Oct 1998) to the EMBL/GenBank/DDBJ databases.
64 EMBL: AC005170; AAC63679.1;
65 FTM: F00292; G196_HY22_28; 1;
66 SP09MVF 466 AA; 50396 MW; 27931AA CEC32;
67
68 Query Match
69 Best local similarity 22.4%; Score 75; DB 19; Length 466;
70 Matches 6; Conservative 11; Mismatches 4; Indels 0; Gaps 0;
71
72 DB 210 PVSNNVIVKCVIIAPIDSP 230
73 : : : : : : : : : : : : : : : :
74 20 SVVQKKEINISVPIIDNS 40
75
76 RESULT 11
77 ID 039226 PRELIMINARY: PRT: 2733 AA.
78 AC 039226:
79 01-JAN-1998 (TEMBUREL_05, Created)
80 01-JAN-1998 (TEMBUREL_05, Last sequence update)
81 01-JAN-1998 (TEMBUREL_05, Last annotation update)
82
83 RNA-DIRECTED RNA POLYMERASE.
84 marine hepatitis virus.
85 Viruses: ssRNA positive-strand viruses, no DNA stage: Nidovirales:
86 Coronaviridae: Coronavirinae.
87 [1]
88 SEQUENCE FROM N.A.
89 STRAIN-MV-A59:
90 DEBARK-GOPFARI I., HINGLEY S.T., CHUA M.M., JIANG X., LAI E.,
91 WEISS S.R.
92 Virology 0:0-0(1997).
93 [2]
94 SEQUENCE FROM N.A.
95 STRAIN-MV-A59:
96 WEISS S.R., DEBARK-GOPFARI I., HINGLEY S.T.
97 Submitted (Oct 1997) to the EMBL/GenBank/DDBJ databases.

```

```

DB EMBL: AF029248; AAB86818.1; -
FW RNA directed RNA polymerase.
82 SEQUENCE 2733 AA; 309126 MW; 09E1E333 CEC32;
83
84 Query Match
85 Best local similarity 22.4%; Score 75; DB 14; Length 2733;
86 Matches 11; Conservative 14; Mismatches 14; Indels 3; Gaps 3;
87
88 DB 908 DYSVILSTGKRGTEDEYF-KNYLR-SAVLOSAGVCVC 947
89 : : : : : : : : : : : : : : : :
90 3 KNYVR-STGSIIFNENAYVGGKFGINTSVLPIDNVCV 43
91
92 RESULT 12
93 ID 011434 PRELIMINARY: PRT: 153 AA.
94 AC 011434:
95 01-JUL-1997 (TEMBUREL_04, Created)
96 01-JAN-1998 (TEMBUREL_05, Last sequence update)
97 01-JAN-1998 (TEMBUREL_05, Last annotation update)
98
99 HYPOTHETICAL 16.3 KD PROTEIN.
100 Viruses: dsDNA viruses, no RNA stage: Adenoviridae: Adenovirinae.
101 [1]
102 SEQUENCE FROM N.A.
103 STRAIN-127:
104 HESS M., BLOECKER H., BRANDI P.
105 "The complete nucleotide sequence of the pig foot and ear virus: an
106 intermediate between mastadenovirus and adenovirus."
107 Virology 238:115-116(1997).
108 [2]
109 SEQUENCE FROM N.A.
110 STRAIN-127:
111 HESS M.
112 Submitted (Nov 1997) to the EMBL/GenBank/DDBJ databases.
113 EMBL: Y09598; CAAT0795.1; -
114 Hypothetical protein
115 G196MVF 153 AA; 16291 MW; DBA08A73 CEC32;
116
117 Query Match
118 Best local similarity 21.8%; Score 73; DB 14; Length 153;
119 Matches 10; Conservative 9; Mismatches 6; Indels 2; Gaps 1;
120
121 DB 78 YTHRYVGS--PGSIAPPPGNIIL 102
122 : : : : : : : : : : : : : : : :
123 16 FNEASYVGGKFGINTSVLPIDNVCV 42
124
125 RESULT 13
126 ID 026982 PRELIMINARY: PRT: 192 AA.
127 AC 026982:
128 01-NOV-1996 (TEMBUREL_01, Created)
129 01-NOV-1996 (TEMBUREL_01, Last sequence update)
130 01-NOV-1999 (TEMBUREL_12, Last annotation update)
131
132 E SELECTIN.
133 Sus scrofa (pig).
134 Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:
135 Eutheria: Cetartiodactyla: Suina: Suidae: Sus.
136 [1]
137 SEQUENCE FROM N.A.
138 STRAIN-127:
139 WINKLER H., KRISTIAN C., CSIZMADIA V., NAFARJAN C., ANRAFHERK J.,
140 KACH F.H.
141 "The intron-exon structure of the porcine E-selectin encoding gene."
142 Gene 176:67-72(1996).
143
144 EMBL: U37521; AAC48680.1; -
145 HSSP: P16581; HESL.
146 PROSITE: PS00615; C_TYPE_SELECTIN_1; 1.
147 PFAM: PF00008; EGF_1.
148 PFAM: PF00059; Lectin_C_1.
149 PFAM: PF00044; Sushi_4.
150 PRINTS: PR00143; SELECTIN.
151 InterPro: Glycoprotein.

```


07-OCT-1997; 018294.
 13-MAR-1997; US-818112.
 11-OCT-1996; US-730510.
 (CORI.) CORIXA CORP.
 Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
 Reed SC, Skolky YAM, Iwakura Y, Vedvick TS;
 WPI: 96-261042/21.
 N-PSDB; V44524.
 Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA used
 to develop products for the detection of M. tuberculosis infection
 and for diagnosis, treatment and prevention of tuberculosis
 Example 30: Page 166, 230pp; English.
 This sequence represents an immunogenic portion of a soluble
 Mycobacterium tuberculosis (M) antigen which can be used in a method
 for inducing protective immunity against tuberculosis (TB). This sequence
 can be formulated into vaccines and/or pharmaceutical compositions for
 immunising against M. tuberculosis infection or may be used for the
 diagnosis of tuberculosis.
 Sequence 284 AA;

Query Match 19.8% Score 75; DB 1; Length 284;
 Best local Similarity 31.3%; Pred. No. 4,366+01;
 Matches 10; Conservative 12; Mismatches 8; Indels 2; Gaps 2;
 64 VEELEPAVIVIKDGLTVMKVSMPROP 95
 24 IETIFADIVYN-SMUS-ISMSPVMPPEP 54

RESULT 3
 W64366 standard; Protein: 284 AA.
 AC W64366;
 09-NOV-1998 (first entry)
 Mycobacterium tuberculosis polypeptide.
 KM Tuberculosis; infection; diagnosis; antigen; XP14.
 CS Mycobacterium tuberculosis strain Erdman.
 FN W09816645-A2.
 23-APR-1998.
 07-OCT-1997; 018214.
 13-MAR-1997; US-818111.
 11-OCT-1996; US-729622.
 (CORI.) CORIXA CORP.
 Campos-Neto A, Dillon DC, Houghton R, Lodes MJ,
 Reed SC, Skolky YAM, Iwakura Y, Vedvick TS;
 WPI: 96-261042/21.
 N-PSDB; V44415.
 New isolated Mycobacterium tuberculosis polypeptides and DNA - used
 to develop products for the detection of M. tuberculosis infection
 and diagnosis of tuberculosis
 Example 3: Page 176-178; 250pp; English.
 This polypeptide is encoded by a DNA molecule (see V44415) isolated
 from a Mycobacterium tuberculosis strain Erdman genome DNA
 expression library using sera from patients having extrapulmonary
 tuberculosis. The invention relates to compositions and methods
 for diagnosing tuberculosis. It provides polypeptides and methods
 W64291 W64379) comprising an antigenic portion of a soluble
 M. tuberculosis antigen, or an immunogenic portion of a M.
 tuberculosis antigen, as well as DNA sequences encoding such
 polypeptides, recombinant expression vectors and transformed or
 transfected host cells. Also claimed are methods and diagnostic
 kits for detecting M. tuberculosis infection in a patient using
 those polypeptides, antibodies or oligonucleotide probes and
 primers.
 Sequence 284 AA;

Query Match 18.8% Score 75; DB 1; Length 284;
 Best local Similarity 31.3%; Pred. No. 4,366+01;
 Matches 10; Conservative 12; Mismatches 8; Indels 2; Gaps 2;
 64 VEELEPAVIVIKDGLTVMKVSMPROP 95
 24 IETIFADIVYN-SMUS-ISMSPVMPPEP 54

RESULT 4
 W18664 standard; Protein: 386 AA.
 AC W18664;
 24-JUL-1997 (first entry)
 Fragmented human NF-H gene (+) frameshift mutant product.
 KM Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;
 KM detection; antibody; probe; cancer; neoplasia; neurodegenerative;
 KM Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
 KM Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
 KM PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
 KM cardiovascular; rheumatoid arthritis.
 CS Homo sapiens.
 FH Key.
 FT misc_difference 1..387
 129..145
 peptide
 Note: "X corresponds to a stop codon in the
 accompanying DNA file, T69796"

W: 772752 A2.
 10 APR-1997.
 02-OCT-1996; IB1106.
 02-OCT-1995; GR-020080.
 11 JAN 1996; US 009832.
 (ROYA-) ROYAL NETHERLANDS ACADEMIS & SCI.
 (OYUT-) UNIV STATE UTRECHT.
 Burbach JPH, Grosveld FG, Van Leeuwen FW;
 WPI: 97-226235/20.
 R-PSDB; 169796.
 Use of mutant genes having frameshift mutation(s) - for developing
 probes for the diagnosis, prevention and treatment of associated
 PT diseases, e.g. cancer or neurodegenerative disease
 BS claim 22; Fig 9; 123pp; English.
 W18663 and W18664 are +2 and +1 frameshift mutations, respectively,
 of a sequence comprising fragments of the coding sequence of the
 human neurofilament subunit NF-H gene. This region contains GAGA motifs.
 CC 1-1162 of the wild-type NF-H gene. This region contains GAGA motifs.
 CC Frameshift mutants of the tau, ubiquitin, apolipoprotein E,
 CC microtubule-associated protein 2 (MAP-2), neurofilament subunit L, M
 CC and H and amyloid A4 genes are claimed. All these genes share a common
 CC GAGAN motif (N-A, G, C or T), which is the site of common GA
 CC dinucleotide deletion(s) that cause neurodegenerative disorders.
 CC Antigenic peptides used for the production of antibodies, and small
 CC nucleic acid sequences derived from frameshift mutants are used in the
 CC diagnosis, prevention and treatment of cancer and neurodegenerative
 CC diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's
 CC syndrome, frontal lobe dementia (Pick's disease), progressive
 CC supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's
 CC disease, multiple sclerosis, and other degenerative diseases such as
 CC cardiovascular disease and rheumatoid arthritis.
 Sequence 386 AA;

Query Match 18.8% Score 75; DB 1; Length 386;
 Best local Similarity 41.2%; Pred. No. 4,460+01;
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

148 EEEVLEAPLPWRRRL 164
 48 QLSMSMSPVMPPEP 54
 RESULT 5
 ID R21200 standard; Protein: 84 AA.
 AC R21200;
 17-JUN-1992 (first entry)
 Human parathyroid hormone analogues [peuA1a]b[hp]H.
 3E PTH; agonist; antagonist; receptor binding.
 OS Synthetic.
 FN W09200753-A.
 24 JAN 1992.
 15-JUL-1990; 004971.

PA 13-JUL-1990; US-553760.
 PA (HEC) Univ of California.
 PI Cohen PA, Nissenson KA, Strowler CJ;
 PI WPI: 92-056643/07
 PI New parathyroid hormone analogues - useful in treating cancer,
 PI osteoporosis, hypercalcemia and hyper-parathyroid conditions
 PS Claim 15; Page 64; 66pp; English.
 CC Residues 35-84 may be absent. The C-terminal gp. may be -COOH,
 CC -COO-M (M = cation), or -(C=O)NH2. 140 specific peptides derived
 CC from human PTH are given in the specification (R21159-257, R21226-
 CC 3250 + R21522-529). Corresp. peptides created using residues 7-84
 CC of bovine and porcine PTH are also claimed. All have mutations at
 CC positions 4, and/or 6, and/or 9 which result in surface side chains
 CC which are useful to modulate receptor binding and activity. They
 CC are useful as agonists and antagonists in the treatment of condi-
 CC tions or diseases involving pit. The peptides are pref. prepd. by
 CC solid phase synthesis. See also R21257 (human genetic), R21258
 CC (bovine genetic) and R21259 (porcine genetic).
 SO Sequence 84 AA;

Query Match 17.8% Score 71; DB 1; Length 84;
 Best Local Similarity 30.0% Prod. No. 9,430-01;
 Matches 9; Conservative 12; Mismatches 7; Indels 2; Gaps 2;

ID 1 SVLEIAMIINIGKHN SMER-VEMLKKL 28
 YF 25 GILFARIVGNSMGI SMISQVWKKRR 54

RESULT 6
 ID R21214 standard; Protein: 84 AA
 AC R21214;
 DI 17-JUN-1992 (first entry)
 DE Human parathyroid hormone analogue, [LeuAla6]hPTH.
 KM PTH agonist; antagonist; receptor binding.
 OS Synthesis.
 PN W09200753-A.
 PD 24-JAN-1992.
 PR 13-JUL-1990; US-553760.
 PA (HEC) Univ of California.
 PI Cohen PA, Nissenson KA, Strowler CJ;
 PI WPI: 92-056643/07.
 PI New parathyroid hormone analogues - useful in treating cancer,
 PI osteoporosis, hypercalcemia and hyper-parathyroid conditions.
 PS Claim 15; Page 65; 66pp; English.
 CC Residues 35-84 may be absent. The C-terminal gp. may be -COOH,
 CC -COO-M (M = cation), or -(C=O)NH2. 140 specific peptides derived
 CC from human PTH are given in the specification (R21159-257, R21226-
 CC 3250 + R21522-529). Corresp. peptides created using residues 7-84
 CC of bovine and porcine PTH are also claimed. All have mutations at
 CC positions 4, and/or 6, and/or 9 which result in surface side chains
 CC which are useful to modulate receptor binding and activity. They
 CC are useful as agonists and antagonists in the treatment of condi-
 CC tions or diseases involving PTH. The peptides are pref. prepd. by
 CC solid phase synthesis. See also R21257 (human genetic), R21258
 CC (bovine genetic) and R21259 (porcine genetic).
 SO Sequence 84 AA;

Query Match 17.8% Score 71; DB 1; Length 84;
 Best Local Similarity 30.0% Prod. No. 9,430-01;
 Matches 9; Conservative 12; Mismatches 7; Indels 2; Gaps 2;

ID 1 SVLEIAMIINIGKHN SMER-VEMLKKL 28
 YF 25 GILFARIVGNSMGI SMISQVWKKRR 54

RESULT 7
 ID P61318 standard; Protein: 319 AA.
 AC P61318;
 DI 28-OCT-1991 (first entry)
 DE Serfatia phospholipase protein.

KM Phospholipids; fatty acids; EN121.
 OS Serfatia sp.
 PN W08606743-A.
 PD 20-NOV-1985.
 PE 09-MAY-1985; DK00051.
 PR 23-DEC-1985; DK-002100.
 PR 09-MAY-1986; WO-DR00041.
 PR 09-JAN-1987; DK-000073.
 PA (HEC) SERFATIA AS.
 PI Molin S, Molin S.
 PI Molin S, Givskov M, Risse F.
 PI WPI: 86-318851/48.
 DE N-PS08; N60877.
 PE Producing bacterial enzyme esp. phospholipase F
 PI which contains INA from strain sp. encoding ext
 PI enzyme.
 PS Disclosure: Fig 9; 64pp; English.
 CC The gene product may be expressed by an E. coli host
 CC a plasmid encoding the phospholipase gene esp. the
 CC Serfatia regulatory sequence. The phospholipase F
 CC be used to hydrolyse fatty acids from phospholipids
 SO Sequence 319 AA;

Query Match 17.8% Score 71; DB 1; Length
 Best Local Similarity 30.0% Prod. No. 9,430-01;
 Matches 9; Conservative 12; Mismatches 7;

ID 131 P0A0-IYSNKKYVLAFAIINWE 17
 YF 27 LKAYCYGRGRLKLEKREKVVW 53

RESULT 8
 ID W20728 standard; Protein: 319 AA.
 AC W20728;
 DI 15-JUL-1997 (first entry)
 DE H. pylori cytoplasmic protein, [Gly105]Hpy14.
 KM Vaccine, prevention; treatment; infection; therapy
 KM binding compound; bacterium; the cyclic activator
 KM inhibitor; duodenal ulcers; chronic gastritis
 KM cytoplasmic mRNA; translation; ribosome; biosynthesis
 OS Helicobacter pylori.
 PN W05640893-A1.
 PD 19-DEC-1996.
 PR 06-JUN-1996; 009122.
 PR 07-JUN-1996; US-487032.
 PR 01-APR-1996; US 6,404,000.
 PA (ASTR) ASTRA AB.
 PI Berglund OT, Smith JS, Molander H.
 PI WPI: 97-052305/05.
 DE N-PS08; 167981.
 PE Helicobacter pylori nucleic acid sequences and for
 PE polypeptides) useful for vaccines to treat or
 PE infection, and to detect Helicobacter.
 PS Claim 61; Page 114; 144pp; English.
 CC The present sequence is a helicobacter pylori cyto-
 CC involved in mRNA translation and ribosome biosynthesis
 CC The protein may be used in a vaccine to prevent or
 CC H. pylori infection or to identify H. pylori poly-
 CC compounds, useful as potential H. pylori life cycle
 CC inhibitors. The genomic sequence of H. pylori (ATC-
 CC determined from overlapping clones generated by re-
 CC sequencing the bacterial DNA. The sequences were ana-
 CC lyzed by computer evaluation. To identify likely H. pylori
 CC vaccine development. The amino acid sequences pred-
 CC various ORF were analysed for significant homology
 CC or exported membrane proteins. Having identified
 CC the sequences of interest, particular regions can be
 CC H. pylori by PCR amplification for recombinant poly-
 CC production, e.g. in E. coli host.
 SO Sequence 319 AA;

FT misc_difference 1603, 1612
 FI misc_difference /Abol: misc feature
 FI misc_difference 2521, 2545 /Abol: misc feature
 PN W09512661-A.
 PD 11-MAY-1995.
 PR 28-OCT-1994: 012423.
 PR 02-NOV-1993: US-148142.
 PA (MERR) MERRICK & CO INC.
 PI Couper M, DAVIS CW, MERRICK & CO INC. Word4 0C, Ymbd:scck J.
 PI Reeves CD, Vliet VA.
 PR 95-193816/25.
 PR N-PSDB: 092323.
 PI Novel DNA encoding of fetal poly-ketide synthase, used to isolate and identify homologues of fetal poly-ketide synthase, and in the treatment of hyper-cholesterolaemia
 PS Claim 12: Figure 2: 10pp; English.
 CC The full-length TKS-encoding DNA in plasmid pLOA was designated pTKS100. Splicing of the introns from the DNA sequence and translation of the 914 nt ORF results in a protein of 3038 AAs (p74171) with a mol. wt. of 269,090 daltons. Inspection of the TKS AA sequence for active site residues and motifs known to be associated with polyketide synthases and fatty acid synthase (FAS) activities resulted in the identification of candidates for expected sites (see FT). Except for the presence of a methyl transferase, the TKS protein is the same as that observed for the fat TKS protein.
 SQ Sequence 4038 AAs

Query Match 17.8% Score 71; DB 1; Length 4038;
 Best Local Similarity 29.5% Prod. No. 9,43601;
 Matches 13; Conservation 11; Mismatches 18; Indels 2; Gaps 2;
 DB 444 LIEVYMEPEQNLRYNNEDCEPMTGVSLPLYLSNKSQSLKI 487
 YQ 2 VIEGYL-S-FNNSRYVHYHNSATVIFAVLYGNSNLSLSM 48

RESULT 12
 ID R23358 standard: Protein: 84 AA.
 AC R23358;
 DT 22-JUN-1992 (first entry)
 DE bovine parathyroid hormone analogue, [Ser110uVala6]bPTH.
 KW PTH; agonist; antagonist; receptor binding.
 OS Synthetic.
 PN W09200753-A.
 PD 23-JAN-1992.
 PR 15-JUL-1990: 004971.
 PR 13-JUL-1990: US-553760.
 PA (REGC) Univ of California.
 PI Cohen FA, Nissenson RA, Strowler GJ.
 PR WPI: 92-056643/07.
 PI New parathyroid hormone analogues - useful in treating cancer, osteoporosis, hypercalcaemia and hyper-parathyroid conditions.
 PS Claim 15: page 65: 86pp; English.
 CC Residues 45-84 may be absent. The C terminal 49, may be CQH, -CQH-4M (M = cation), or -(P-O)NH2. 112 specific peptides derived from bovine PTH are given in the specification (R23251-482, and R23540-549). Corresp. peptides created using residues 7-94 of human and porcine PTH are also claimed. All have mutations at positions 3, and/or 6, and/or 9 which result in surface side chains which are useful to modulate receptor binding and activity. They are useful as agonists and antagonists in the treatment of conditions or diseases involving PTH. The peptides are prod. by solid phase synthesis. See also R21257 (human genetic), R21258 (bovine genetic) and R21259 (porcine genetic).
 SQ Sequence 84 AAs

Query Match 17.6% Score 70; DB 1; Length 84;
 Best Local Similarity 43.8% Prod. No. 1,14602;
 Matches 10; Conservation 11; Mismatches 7; Indels 2; Gaps 2;

ID 1 SVLEIAPMNDKEHNSGPE VFWLKKL 26
 YQ 25 GILFADQYGNSSNGLSMDLSEVFWWRRP 54

RESULT 13
 ID R23344 standard: Protein: 84 AA
 AC R23344;
 DT 22-JUN-1992 (first entry)
 DE bovine parathyroid hormone analogue, [Ser110uVala6]bPTH.
 KW PTH; agonist; antagonist; receptor binding.
 OS Synthetic.
 PN W09200753-A.
 PD 23-JAN-1992.
 PR 15-JUL-1990: 004971.
 PR 13-JUL-1990: US-553760.
 PA (REGC) Univ of California.
 PI Cohen FA, Nissenson RA, Strowler GJ.
 PR WPI: 92-056643/07.
 PI New parathyroid hormone analogues - useful in treating osteoporosis, hypercalcaemia and hyper-parathyroid conditions.
 PS Claim 15: page 64: 86pp; English.
 CC Residues 35-84 may be absent. The C-terminal 49, may be CQH-4M (M = cation), or -(P-O)NH2. 112 specific peptides derived from bovine PTH are given in the specification (R23251-482, and R23540-549). Corresp. peptides created using positions 3, and/or 6, and/or 9 which result in surface side chains which are useful to modulate receptor binding and activity. They are useful as agonists and antagonists in the treatment of conditions or diseases involving PTH. The peptides are prod. by solid phase synthesis. See also R21257 (human genetic), R21258 (bovine genetic) and R21259 (porcine genetic).
 SQ Sequence 84 AAs

Query Match 17.7% Score 70; DB 1; Length 84;
 Best Local Similarity 43.8% Prod. No. 1,14602;
 Matches 10; Conservation 11; Mismatches 7;

DB 1 SVLEIAPMNDKEHNSGPE VFWLKKL 26
 YQ 25 GILFADQYGNSSNGLSMDLSEVFWWRRP 54
 RESULT 14
 ID W44269 standard: Protein: 84 AA
 AC W44269;
 DT 19-MAY-1998 (first entry)
 DE Hybrid DNA protein comprising a family 45 cellulase, cellulase core region, located on microorganism 11
 KW hybrid DNA.
 OS Chimeric - Humicola insolens.
 OS Chimeric - Humicola insolens.
 PN W09743409-A2.
 PD 20-NOV-1997.
 PR 12-MAY-1997: PK0216.
 PR 10-MAY-1996: PK-000562.
 PA (HVA) H V - HVA-28 AS.
 PI Dalgaard H, Dalgaard H, Koefgen S, Sander H.
 PR WPI: 98-03686/01.
 PI Isolation novel DNA sequences from microorganisms for culturing the microorganism.
 PS Example 4: page 52-53: 74pp; English.
 CC The present sequence represents a protein encoded from an example of the present invention. The present describes a novel method for providing a novel DNA polypeptide from a microorganism with an activity method comprises: (1) for amplification of the DNA homology to (11) known sequence(s) encoding a polypeptide of interest; (11) linking the amplified PCR product at sequence and a 5' structure to non-sequence (11);

 RELEASE 3.1A John F. Collins, Bioinformatics Research Unit,
 Copyright (C) 1994-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

(TM)

Release 3.1A John F. Collins, Bioinformatics Research Unit,
 Copyright (C) 1994-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mustch_FP Protein - Protein database search, using Smith-Waterman algorithm
 Run on: Fri Aug 25 18:09:28 2000: Maxpat time 2.62 seconds
 Labelled output not generated. 416,225 William coll updates/sec

Title: 20S-09-257 585-8
 Description: (1-54) from US09257585.pcp
 Perfect Score: 498
 Sequence: 1 MVEEYLSRNSSPV/NVHCP NSNLSLSMSTVENPPPI 54

Scoring table:
 PAM 150
 Gap 11

Search: 152433 seqs, 15129240 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: a-issind
 1:5A_QOMH 2:5H_QOMH 3:6_QOMH 4:PCL_QOMH 5:backfiles1

Statistics: Mean 24.818: Variance 91.496: scale 0.272

Prod. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description	Prod. No.
1	75	18.8	188	2 US-08-726- Sequence 57, Applicat	2.16e+01
2	71	17.8	203	4 US-08-640- Sequence 18, Applicat	4.77e+01
3	71	17.8	308	2 US-08-637- Sequence 2, Applicat	4.77e+01
4	71	17.8	408	1 US-08-450- Sequence 2, Applicat	4.77e+01
5	70	17.3	360	2 US-08-483- Sequence 17, Applicat	5.80e+01
6	69	17.3	344	2 US-08-257- Sequence 9, Applicat	7.04e+01
7	69	17.3	444	4 US-08-477- Sequence 9, Applicat	7.04e+01
8	69	17.3	345	4 US-08-866- Sequence 1, Applicat	7.04e+01
9	69	17.3	544	4 US-08-866- Sequence 5, Applicat	7.04e+01
10	69	17.3	545	1 US-07-906- Sequence 10, Applicat	7.04e+01
11	69	17.3	545	2 US-08-167- Sequence 10, Applicat	7.04e+01
12	69	17.3	545	2 US-08-579- Sequence 10, Applicat	7.04e+01
13	69	17.3	545	1 US-08-208- Sequence 10, Applicat	7.04e+01
14	69	17.3	650	1 US-08-121- Sequence 60, Applicat	7.04e+01
15	69	17.3	650	4 US-08-833- Sequence 60, Applicat	7.04e+01
16	69	17.3	650	1 US-08-835- Sequence 60, Applicat	7.04e+01
17	69	17.3	650	4 PCL-0894-1 Sequence 60, Applicat	7.04e+01
18	69	17.3	650	2 US-08-670- Sequence 60, Applicat	7.04e+01
19	68	17.1	412	1 US-08-708- Sequence 12, Applicat	8.54e+01
20	68	17.1	550	1 US-08-674- Sequence 29, Applicat	8.54e+01
21	68	17.1	711	1 US-08-772- Sequence 12, Applicat	8.54e+01
22	67	16.8	422	1 US-08-014- Sequence 2, Applicat	1.03e+02
23	67	16.8	372	1 US-08-456- Sequence 3, Applicat	1.03e+02

24	67	16.8	422	2 US-08-486- Sequence 3, Applicat	1.03e+02
25	67	16.8	422	1 US-08-470- Sequence 3, Applicat	1.03e+02
26	67	16.8	692	1 US-08-965- Sequence 2, Applicat	1.03e+02
27	67	16.8	692	1 US-08-291- Sequence 2, Applicat	1.03e+02
28	67	16.8	692	2 US-08-485- Sequence 2, Applicat	1.03e+02
29	67	16.8	692	2 US-08-475- Sequence 2, Applicat	1.03e+02
30	66	16.6	66	1 US-07-966- Sequence 24, Applicat	1.03e+02
31	66	16.6	86	1 US-08-476- Sequence 1, Applicat	1.03e+02
32	66	16.6	86	1 US-08-485- Sequence 1, Applicat	1.03e+02
33	66	16.6	86	1 US-07-966- Sequence 1, Applicat	1.03e+02
34	66	16.6	86	1 US-08-485- Sequence 24, Applicat	1.03e+02
35	66	16.6	86	1 US-08-475- Sequence 24, Applicat	1.03e+02
36	66	16.6	86	2 US-08-475- Sequence 24, Applicat	1.03e+02
37	66	16.6	95	1 US-08-476- Sequence 2, Applicat	1.03e+02
38	66	16.6	95	1 US-07-966- Sequence 2, Applicat	1.03e+02
39	66	16.6	95	2 US-08-475- Sequence 2, Applicat	1.03e+02
40	66	16.6	95	1 US-08-485- Sequence 2, Applicat	1.03e+02
41	66	16.6	264	4 US-09-411-1 Atrial No. 1	1.03e+02
42	66	16.6	549	3 US-08-651- Sequence 3, Applicat	1.03e+02
43	66	16.6	549	4 US-08-772- Sequence 3, Applicat	1.03e+02
44	66	16.6	549	4 US-08-746- Sequence 3, Applicat	1.03e+02
45	65	16.3	1749	2 US-08-470- Sequence 7, Applicat	1.03e+02

ALIGNMENTS

RESULT 1
 ID US-08-726-306A-57 P1: 148 AA
 AC xxxxxx
 XX

Sequence 57, Application US/08726306A

Sequence 57, Application US/08726306A

Patent No. 5958684

GENERAL INFORMATION:

APPLICANT: van Leeuwen, Frederik Willem
 APPLICANT: Grosveld, Franklin G.
 TITLE OF INVENTION: HLA-RELATED METHOD AND REAGENTS
 NUMBER OF SEQUENCES: 199
 CORRESPONDENCE ADDRESS:

ADDRESS: Biogen & Watson, Ltd.
 STREET: 1 Finsbury Pavement
 CITY: Boston
 STATE: MA
 COUNTRY: US

ZIP: 02111

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 5.25 inch, 1.44 MB
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC DOS/MS-DOS

SOFTWARE: WORDPROCESS 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08726306A

FILING DATE: 02/06/1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08726306A

FILING DATE: 02/06/1996

ATTORNEY/AGENT INFORMATION:

NAME: Williams, Patricia Kathleen M.

REGISTRATION NUMBER: 44,480

TELEPHONE: (617) 845-0100

TELEFAX: (617) 845-0100

INFORMATION FOR SEQ. ID NO. 57

SEQUENCE CHARACTERISTICS:

TELEFAX: (212) 969-9741/8864
TELEPHONE: (212) 969-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SUB ID NO: 97

CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein

FEATURE:
OTHER INFORMATION: HLM domain of GRB-7
SEQUENCE 335 AA: 47724 MW: 569630 CN:

Query Match 17.4% Score 69 DB 3 Length 335
Best Local Similarity 42.0% Prod No. 7 040001
Matches 8: Conservative 0 Mismatches 7 Indels 1 Gaps 1

DB 291 ADEVITPPPALSAAMEEAAMKEE 415
11: 11 11 11
29 ADEVITPPPALSAAMEEAAMKEE 52

RESULT 9
ID US-08-066-381A-5 STANDARD: PRT: 534 AA.

AC xxxxxx

DE Sequence 9, Application US/08066381A

XX Sequence 9, Application US/08066381A
XX Patent No. 6045797

XX GENERAL INFORMATION:

XX APPLICANT: Ron Lewis Margolis

XX TITLE OF INVENTION: METHODS FOR TREATMENT OF DISEASES

XX TITLE OF INVENTION: OF DISEASES OR CONDITIONS ASSOCIATED

XX NUMBER OF SEQUENCES: 6

XX CORRESPONDENCE ADDRESS:

XX ADDRESS: 1900 S Lyon

XX STREET: 613 West Fifth Street

XX STATE: Suite 4700

XX CITY: Los Angeles

XX STATE: California

XX COUNTRY: U.S.A

XX ZIP: 90071-2066

XX COMPUTER READABLE FORM:

XX MEDIUM TYPE: 3.5" Diskette, 1.44 MB

XX OPERATING SYSTEM: IBM PC DOS 5.0

XX SOFTWARE: SYSTEM FOR WINDOVS 2.0

XX CURRENT APPLICATION DATA:

XX APPLICATION NUMBER: US/08/066, 381A

XX FILING DATE: May 30, 1997

XX CLASSIFICATION: 530

XX PRIOR APPLICATION DATA:

XX APPLICATION NUMBER: 08/212,234

XX FILING DATE: March 14, 1994

XX APPLICATION NUMBER:

XX FILING DATE:

XX ATTORNEY/AGENT INFORMATION:

XX NAME: Wardlaw, Richard J.

XX REGISTRATION NUMBER: 32,327

XX TELEPHONE: (213) 489-1600

XX TELEPHONE: (213) 489-1600

XX TELEFAX: (213) 955-0440

XX TELEFAX: (213) 955-0440

XX INFORMATION FOR SEQ ID NO: 5:

XX SEQUENCE CHARACTERISTICS:

XX LENGTH: 534 amino acids

XX TYPE: amino acid

XX STRANDEDNESS: single

XX TOPOLOGY: linear

XX MOLECULE TYPE: protein

XX FEATURE:

XX OTHER INFORMATION: GRB-7

XX SEQUENCE: 534 AA: 50827 MW: 142851 CN:

Query Match 17.4% Score 69 DB 3 Length 335
Best Local Similarity 42.0% Prod No. 7 040001
Matches 8: Conservative 0 Mismatches 7

DB 383 ADEVITPPPALSAAMEEAAMKEE 417
11: 11 11 11
29 ADEVITPPPALSAAMEEAAMKEE 52

RESULT 10
ID US 07 906 349A 10 STANDARD: PRT: 534 AA.

AC xxxxxx

DE Sequence 10, Application US/07/096349A

XX Sequence 10, Application US/07/096349A
XX Patent No. 544064

XX GENERAL INFORMATION:

XX APPLICANT: Schlessinger, Joseph

XX APPLICANT: Skolnik, Edward Y.

XX APPLICANT: Margolis, Benjamin L.

XX TITLE OF INVENTION: A METHOD FOR EXPRESSION OF NIN

XX TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS

XX NUMBER OF SEQUENCES: 10

XX CORRESPONDENCE ADDRESS:

XX ADDRESS: Brown and Nodmark

XX STREET: 419 Seventh Street, N.W.

XX CITY: Washington

XX STATE: D.C.

XX COUNTRY: USA

XX ZIP: 20004

XX COMPUTER READABLE FORM:

XX MEDIUM TYPE: 5.25" disk

XX OPERATING SYSTEM: IBM PC compatible

XX SOFTWARE: Patent to Bioscience #10, Ver 3.0 #1

XX CURRENT APPLICATION DATA:

XX APPLICATION NUMBER: 08/077,066, 349A

XX FILING DATE: 10-01-97

XX CLASSIFICATION: 435

XX FRIER APPLICATION DATA:

XX APPLICATION NUMBER: 08/077,066, 349A

XX FILING DATE: 10-01-97

XX TELECOMMUNICATION INFORMATION:

XX TELEPHONE: 202-528-5137

XX TELEFAX: 202-528-5137

XX INFORMATION FOR SEQ ID NO: 10:

XX SEQUENCE CHARACTERISTICS:

XX LENGTH: 535 amino acids

XX TYPE: amino acid

XX STRANDEDNESS: single

XX TOPOLOGY: linear

XX MOLECULE TYPE: protein

XX SEQUENCE: 535 AA: 53553 MW: 141072 CN:

XX Query Match 17.4% Score 69 DB 3 Length 335

XX Best Local Similarity 42.0% Prod No. 7 040001

XX Matches 8: Conservative 0 Mismatches 7

DB 384 ADEVITPPPALSAAMEEAAMKEE 417

11: 11 11 11

29 ADEVITPPPALSAAMEEAAMKEE 52

RESULT 11

ID US-08-167-035-10 STANDARD: PRT: 534 AA.

AC xxxxxx

```

01 XX Sequence 10, Application US/08167035
02 DE
03 XX
04 DE
05 DE Sequence 10, Application US/08167035
06 DE Patent No. 5618691
07 DE
08 DE GENERAL INFORMATION:
09 DE
10 DE APPLICANT: Schlessinger, Joseph
11 DE
12 DE APPLICANT: Skolnick, Edward Y.
13 DE
14 DE APPLICANT: Marzolis, Benjamin L.
15 DE
16 DE TITLE OF INVENTION: NOVEL EXPRESSION SYSTEMS METHOD FOR
17 DE
18 DE TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR ENCAPSULATED CYTOSINE
19 DE
20 DE NUMBER OF SEQUENCES: 50
21 DE
22 DE CORRESPONDENCE ADDRESS:
23 DE
24 DE ADDRESSEE: PENNIE & EDMONDS
25 DE
26 DE STREET: 1155 Avenue of the Americas
27 DE
28 DE CITY: New York
29 DE
30 DE STATE: New York
31 DE
32 DE COUNTRY: 10036-2711
33 DE
34 DE ZIP: 10036-2711
35 DE
36 DE COMPUTER READABLE FORM:
37 DE
38 DE MEDIUM TYPE: Floppy disk
39 DE
40 DE COMPUTER: IBM PC compatible
41 DE
42 DE OPERATING SYSTEM: PC-DOS/MS-DOS
43 DE
44 DE SOFTWARE: Patent Release #1.0, Version #1.30
45 DE
46 DE CURRENT APPLICATION DATA:
47 DE
48 DE APPLICATION NUMBER: US/06/167 035
49 DE
50 DE FILING DATE: 16-DEC-1993
51 DE
52 DE CLASSIFICATION: 435
53 DE
54 DE ATTORNEY/AGENT INFORMATION:
55 DE
56 DE NAME: Coruzzi, Laura A.
57 DE
58 DE REGISTRATION NUMBER: 30,742
59 DE
60 DE REFERENCE/WORKET NUMBER: 7683-062
61 DE
62 DE TELECOMMUNICATION INFORMATION:
63 DE
64 DE TELEPHONE: (212) 790-9090
65 DE
66 DE TELEFAX: (212) 869 9741/8864
67 DE
68 DE TELE: 66141PENNI
69 DE
70 DE INFORMATION FOR SEQ ID NO: 10:
71 DE
72 DE SEQUENCE CHARACTERISTICS:
73 DE
74 DE LENGTH: 535 amino acids
75 DE
76 DE TYPE: amino acid
77 DE
78 DE TOPOLOGY: unknown
79 DE
80 DE MOLECULE TYPE: protein
81 DE
82 DE SEQUENCE 535 AA: 59973 MM: 1418700 CN:
83 DE
84 DE
85 DE
86 DE
87 DE
88 DE
89 DE
90 DE
91 DE
92 DE
93 DE
94 DE
95 DE
96 DE
97 DE
98 DE
99 DE
00 DE
01 DE
02 DE
03 DE
04 DE
05 DE
06 DE
07 DE
08 DE
09 DE
10 DE
11 DE
12 DE
13 DE
14 DE
15 DE
16 DE
17 DE
18 DE
19 DE
20 DE
21 DE
22 DE
23 DE
24 DE
25 DE
26 DE
27 DE
28 DE
29 DE
30 DE
31 DE
32 DE
33 DE
34 DE
35 DE
36 DE
37 DE
38 DE
39 DE
40 DE
41 DE
42 DE
43 DE
44 DE
45 DE
46 DE
47 DE
48 DE
49 DE
50 DE
51 DE
52 DE
53 DE
54 DE
55 DE
56 DE
57 DE
58 DE
59 DE
60 DE
61 DE
62 DE
63 DE
64 DE
65 DE
66 DE
67 DE
68 DE
69 DE
70 DE
71 DE
72 DE
73 DE
74 DE
75 DE
76 DE
77 DE
78 DE
79 DE
80 DE
81 DE
82 DE
83 DE
84 DE
85 DE
86 DE
87 DE
88 DE
89 DE
90 DE
91 DE
92 DE
93 DE
94 DE
95 DE
96 DE
97 DE
98 DE
99 DE
00 DE
01 DE
02 DE
03 DE
04 DE
05 DE
06 DE
07 DE
08 DE
09 DE
10 DE
11 DE
12 DE
13 DE
14 DE
15 DE
16 DE
17 DE
18 DE
19 DE
20 DE
21 DE
22 DE
23 DE
24 DE
25 DE
26 DE
27 DE
28 DE
29 DE
30 DE
31 DE
32 DE
33 DE
34 DE
35 DE
36 DE
37 DE
38 DE
39 DE
40 DE
41 DE
42 DE
43 DE
44 DE
45 DE
46 DE
47 DE
48 DE
49 DE
50 DE
51 DE
52 DE
53 DE
54 DE
55 DE
56 DE
57 DE
58 DE
59 DE
60 DE
61 DE
62 DE
63 DE
64 DE
65 DE
66 DE
67 DE
68 DE
69 DE
70 DE
71 DE
72 DE
73 DE
74 DE
75 DE
76 DE
77 DE
78 DE
79 DE
80 DE
81 DE
82 DE
83 DE
84 DE
85 DE
86 DE
87 DE
88 DE
89 DE
90 DE
91 DE
92 DE
93 DE
94 DE
95 DE
96 DE
97 DE
98 DE
99 DE
00 DE
01 DE
02 DE
03 DE
04 DE
05 DE
06 DE
07 DE
08 DE
09 DE
10 DE
11 DE
12 DE
13 DE
14 DE
15 DE
16 DE
17 DE
18 DE
19 DE
20 DE
21 DE
22 DE
23 DE
24 DE
25 DE
26 DE
27 DE
28 DE
29 DE
30 DE
31 DE
32 DE
33 DE
34 DE
35 DE
36 DE
37 DE
38 DE
39 DE
40 DE
41 DE
42 DE
43 DE
44 DE
45 DE
46 DE
47 DE
48 DE
49 DE
50 DE
51 DE
52 DE
53 DE
54 DE
55 DE
56 DE
57 DE
58 DE
59 DE
60 DE
61 DE
62 DE
63 DE
64 DE
65 DE
66 DE
67 DE
68 DE
69 DE
70 DE
71 DE
72 DE
73 DE
74 DE
75 DE
76 DE
77 DE
78 DE
79 DE
80 DE
81 DE
82 DE
83 DE
84 DE
85 DE
86 DE
87 DE
88 DE
89 DE
90 DE
91 DE
92 DE
93 DE
94 DE
95 DE
96 DE
97 DE
98 DE
99 DE
00 DE
01 DE
02 DE
03 DE
04 DE
05 DE
06 DE
07 DE
08 DE
09 DE
10 DE
11 DE
12 DE
13 DE
14 DE
15 DE
16 DE
17 DE
18 DE
19 DE
20 DE
21 DE
22 DE
23 DE
24 DE
25 DE
26 DE
27 DE
28 DE
29 DE
30 DE
31 DE
32 DE
33 DE
34 DE
35 DE
36 DE
37 DE
38 DE
39 DE
40 DE
41 DE
42 DE
43 DE
44 DE
45 DE
46 DE
47 DE
48 DE
49 DE
50 DE
51 DE
52 DE
53 DE
54 DE
55 DE
56 DE
57 DE
58 DE
59 DE
60 DE
61 DE
62 DE
63 DE
64 DE
65 DE
66 DE
67 DE
68 DE
69 DE
70 DE
71 DE
72 DE
73 DE
74 DE
75 DE
76 DE
77 DE
78 DE
79 DE
80 DE
81 DE
82 DE
83 DE
84 DE
85 DE
86 DE
87 DE
88 DE
89 DE
90 DE
91 DE
92 DE
93 DE
94 DE
95 DE
96 DE
97 DE
98 DE
99 DE
00 DE
01 DE
02 DE
03 DE
04 DE
05 DE
06 DE
07 DE
08 DE
09 DE
10 DE
11 DE
12 DE
13 DE
14 DE
15 DE
16 DE
17 DE
18 DE
19 DE
20 DE
21 DE
22 DE
23 DE
24 DE
25 DE
26 DE
27 DE
28 DE
29 DE
30 DE
31 DE
32 DE
33 DE
34 DE
35 DE
36 DE
37 DE
38 DE
39 DE
40 DE
41 DE
42 DE
43 DE
44 DE
45 DE
46 DE
47 DE
48 DE
49 DE
50 DE
51 DE
52 DE
53 DE
54 DE
55 DE
56 DE
57 DE
58 DE
59 DE
60 DE
61 DE
62 DE
63 DE
64 DE
65 DE
66 DE
67 DE
68 DE
69 DE
70 DE
71 DE
72 DE
73 DE
74 DE
75 DE
76 DE
77 DE
78 DE
79 DE
80 DE
81 DE
82 DE
83 DE
84 DE
85 DE
86 DE
87 DE
88 DE
89 DE
90 DE
91 DE
92 DE
93 DE
94 DE
95 DE
96 DE
97 DE
98 DE
99 DE
00 DE
01 DE
02 DE
03 DE
04 DE
05 DE
06 DE
07 DE
08 DE
09 DE
10 DE
11 DE
12 DE
13 DE
14 DE
15 DE
16 DE
17 DE
18 DE
19 DE
20 DE
21 DE
22 DE
23 DE
24 DE
25 DE
26 DE
27 DE
28 DE
29 DE
30 DE
31 DE
32 DE
33 DE
34 DE
35 DE
36 DE
37 DE
38 DE
39 DE
40 DE
41 DE
42 DE
43 DE
44 DE
45 DE
46 DE
47 DE
48 DE
49 DE
50 DE
51 DE
52 DE
53 DE
54 DE
55 DE
56 DE
57 DE
58 DE
59 DE
60 DE
61 DE
62 DE
63 DE
64 DE
65 DE
66 DE
67 DE
68 DE
69 DE
70 DE
71 DE
72 DE
73 DE
74 DE
75 DE
76 DE
77 DE
78 DE
79 DE
80 DE
81 DE
82 DE
83 DE
84 DE
85 DE
86 DE
87 DE
88 DE
89 DE
90 DE
91 DE
92 DE
93 DE
94 DE
95 DE
96 DE
97 DE
98 DE
99 DE
00 DE
01 DE
02 DE
03 DE
04 DE
05 DE
06 DE
07 DE
08 DE
09 DE
10 DE
11 DE
12 DE
13 DE
14 DE
15 DE
16 DE
17 DE
18 DE
19 DE
20 DE
21 DE
22 DE
23 DE
24 DE
25 DE
26 DE
27 DE
28 DE
29 DE
30 DE
31 DE
32 DE
33 DE
34 DE
35 DE
36 DE
37 DE
38 DE
39 DE
40 DE
41 DE
42 DE
43 DE
44 DE
45 DE
46 DE
47 DE
48 DE
49 DE
50 DE
51 DE
52 DE
53 DE
54 DE
55 DE
56 DE
57 DE
58 DE
59 DE
60 DE
61 DE
62 DE
63 DE
64 DE
65 DE
66 DE
67 DE
68 DE
69 DE
70 DE
71 DE
72 DE
73 DE
74 DE
75 DE
76 DE
77 DE
78 DE
79 DE
80 DE
81 DE
82 DE
83 DE
84 DE
85 DE
86 DE
87 DE
88 DE
89 DE
90 DE
91 DE
92 DE
93 DE
94 DE
95 DE
96 DE
97 DE
98 DE
99 DE
00 DE
01 DE
02 DE
03 DE
04 DE
05 DE
06 DE
07 DE
08 DE
09 DE
10 DE
11 DE
12 DE
13 DE
14 DE
15 DE
16 DE
17 DE
18 DE
19 DE
20 DE
21 DE
22 DE
23 DE
24 DE
25 DE
26 DE
27 DE
28 DE
29 DE
30 DE
31 DE
32 DE
33 DE
34 DE
35 DE
36 DE
37 DE
38 DE
39 DE
40 DE
41 DE
42 DE
43 DE
44 DE
45 DE
46 DE
47 DE
48 DE
49 DE
50 DE
51 DE
52 DE
53 DE
54 DE
55 DE
56 DE
57 DE
58 DE
59 DE
60 DE
61 DE
62 DE
63 DE
64 DE
65 DE
66 DE
67 DE
68 DE
69 DE
70 DE
71 DE
72 DE
73 DE
74 DE
75 DE
76 DE
77 DE
78 DE
79 DE
80 DE
81 DE
82 DE
83 DE
84 DE
85 DE
86 DE
87 DE
88 DE
89 DE
90 DE
91 DE
92 DE
93 DE
94 DE
95 DE
96 DE
97 DE
98 DE
99 DE
00 DE
01 DE
02 DE
03 DE
04 DE
05 DE
06 DE
07 DE
08 DE
09 DE
10 DE
11 DE
12 DE
13 DE
14 DE
15 DE
1
```

```

01 NUMBER OF SEQUENCES: 50
02 CORRESPONDENCE ADDRESS:
03 ADDRESSEE: PENNIE & EDMONDS
04 STREET: 1155 Avenue of the Americas
05 CITY: New York
06 STATE: New York
07 COUNTRY: 10036-2711
08 ZIP: 10036-2711
09
10 COMPUTER READABLE FORM:
11
12 MEDIUM TYPE: Floppy disk
13 COMPUTER: IBM PC compatible
14 OPERATING SYSTEM: PC-DOS/MS-DOS
15 SOFTWARE: Patent Release #1.0, Version #1.30
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER: US/08/539, 005
18 FILING DATE: 4-OCT-1995
19 CLASSIFICATION: 435
20
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US 08/167,035
23 FILING DATE: 16-DEC-1993
24 CLASSIFICATION: 435
25
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Coruzzi, Laura A.
28 REGISTRATION NUMBER: 40,742
29 REFERENCE/DOCKET NUMBER: 7683-062
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: (212) 790-9090
32 TELEFAX: (212) 865-9741/8864
33 TELEX: 66141 PENNIE
34 INFORMATION FOR SEQ. ID NO: 10:
35
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 535 amino acids
38 TYPE: amino acid
39 TOPOLOGY: unknown
40 MOLECULE TYPE: protein
41
42 SEQUENCE: 535 AA; 59973 MW; 1438760 CN;
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
9
```


Query Match 17.38; Score 69; PB 3; Length 650;
Best Local Similarly 20.94; Pred. No. 7.04e+01;
Matches 9; Conservative 13; Mismatches 20; Indels 1; Gaps 1;

16 46 TKNQAVCPYDPRHNSLVLADNHLVSGTVA-DHSGSDP1YR 87
::: ||| : : : : : : : : : : : :
17 8 SRNSSVWVNYHQPSSAIGII FADGIYGNSSNQI SI SMDSVVEWP 50

Search completed: Fri Aug 25 18:09:37 2000
Load time : 9 secs.

 MWSEKELI (1M)

Release 3.1A John F. Collins, BioComputing Research Unit,
 Copyright (c) 1994-1998 University of Cambridge, U.K.
 Distribution rights by Oxford Molecular Ltd

Match:pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Aug 25 18:08:01 2000: MasPar Lite 1.72 Seconds
 Label: output not generated. 449,571 Million cell updates/sec

File: >US-09-257-585-8
 Description: (1-54) from US09257585.pep
 Perfect Score: 398
 Sequence: 1 MYRKYLSKNSSRVGVNHC.....NSNQLSLMSIPVKKRRRI 54

Scoring Table:
 PAM 150
 Gap 11

Searched: 85661 seqs, 40989116 residues

Post processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot 48
 TrEMBL 1
 SwissProt 48

Statistics: Mean 36.45%, Variance 52.52%, Scale 0.694

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	94	21.6	496	1	SRM_MOUSE TYROSINE-PROTEIN KINASE	4.786+04
2	83	20.9	156	1	CTIP_SYTR3 CYTOSOLIC TYPING	5.172+02
3	80	20.1	1331	1	CYAB_LITDO KEEFOPOR-TYPE ADENYLAT	1.746+01
4	79	19.8	212	1	LIP_BACSO LIPASE PRECURSOR (EC 3	2.596+01
5	76	19.1	278	1	IEF8_VZVP GENE 63/70 PROTEIN	8.346+01
6	76	19.1	434	1	TEST_RAT TESTIN 1/2 PRECURSOR (8.346+01
7	76	19.1	425	1	FIXO_PCELI FIXO PROTEIN	8.346+01
8	76	19.1	758	1	PCUA_YEMEN FERROCHROME RECEPTOR	8.346+01
9	76	19.1	801	1	VP34_SCHPO PHOSPHATIDYLINOSITOL	8.346+01
10	75	18.8	425	1	GRH2_MYCTO PIRGELAVIN BIOSYNTHET	1.228+00
11	75	18.8	863	1	Y589_CAEEL HYPOPHOSPHATIDYL	1.228+00
12	75	18.8	1258	1	CTIP_FINER PROTEIN 26	1.228+00
13	74	18.6	808	1	GLND_MYCTO GLENDIN-PI 1 UNIDENTI	1.786+00
14	73	18.3	257	1	ABH1_CERCA ALCOHOL DEHYDROGENASE	2.606+00
15	73	18.3	566	1	Y076_YEAST HYDROLYTIC 6A.1 NDA	2.606+00
16	72	18.1	277	1	TPR2_PCEAD TP2 PARCE TYPE 2	3.706+00
17	72	18.1	276	1	TPR2_PCEAD TP2 PARCE TYPE 2	3.706+00
18	72	18.1	412	1	MEAL_SALTY M-ALBOMERIN BIOSYNTH	3.706+00
19	72	18.1	5147	1	FAT1_DROMO CAUDERIN-RELATED TUMOR	3.766+00
20	71	17.8	214	1	VIF_STYRK VARIATION IMPROVING FAC	5.426+00
21	71	17.8	277	1	SEAC1_LVA SEAC1 LVA SEAC1	5.426+00
22	71	17.8	405	1	FRT_BELTU METHYLIN TNA FORMYLT	5.426+00
23	71	17.8	419	1	PAL_SERLI EXTRACELLULAR PHOSPHOL	5.426+00

24	71	17.8	577	1	PA3MA CYTOSOLIC ME	5.426+00
25	71	17.8	579	1	PM1_MESOR 2.3 DISFIB	5.426+00
26	71	17.8	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00
27	70	17.6	577	1	PA3MA CYTOSOLIC ME	5.426+00
28	70	17.6	579	1	PM1_MESOR 2.3 DISFIB	5.426+00
29	70	17.6	577	1	PA3MA CYTOSOLIC ME	5.426+00
30	70	17.6	579	1	PM1_MESOR 2.3 DISFIB	5.426+00
31	70	17.6	577	1	PA3MA CYTOSOLIC ME	5.426+00
32	70	17.6	579	1	PM1_MESOR 2.3 DISFIB	5.426+00
33	70	17.6	577	1	PA3MA CYTOSOLIC ME	5.426+00
34	70	17.6	579	1	PM1_MESOR 2.3 DISFIB	5.426+00
35	70	17.6	577	1	PA3MA CYTOSOLIC ME	5.426+00
36	70	17.6	579	1	PM1_MESOR 2.3 DISFIB	5.426+00
37	69	17.4	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00
38	69	17.4	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00
39	69	17.4	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00
40	69	17.4	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00
41	69	17.4	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00
42	69	17.4	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00
43	69	17.4	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00
44	69	17.4	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00
45	69	17.4	117	1	BMOL_BRAVE BETA-2 MICRO	5.426+00

ALIGNMENTS

RESULT ID	1	STANDARD	PEP	496 AA
1	SRM_MOUSE	Q62370: Q62360:		
2	01-NOV-1997 (Rel. 45, created)			
3	01-NOV-1997 (Rel. 45, last sequence update)			
4	01-NOV-1997 (Rel. 45, last annotation update)			
5	TYROSINE-PROTEIN KINASE SEQ (EC 2.7.1.112) (447/54)			
6	SRM_MOUSE			
7	Mus musculus (Mouse)			
8	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia: Eutheria: Rodentia: Sciurognathi: Murina: [1]			
9	SEQUENCE FROM N.A.			
10	SRM_MOUSE			
11	SRM_MOUSE			
12	SRM_MOUSE			
13	SRM_MOUSE			
14	SRM_MOUSE			
15	SRM_MOUSE			
16	SRM_MOUSE			
17	SRM_MOUSE			
18	SRM_MOUSE			
19	SRM_MOUSE			
20	SRM_MOUSE			
21	SRM_MOUSE			
22	SRM_MOUSE			
23	SRM_MOUSE			
24	SRM_MOUSE			
25	SRM_MOUSE			
26	SRM_MOUSE			
27	SRM_MOUSE			
28	SRM_MOUSE			
29	SRM_MOUSE			
30	SRM_MOUSE			
31	SRM_MOUSE			
32	SRM_MOUSE			
33	SRM_MOUSE			
34	SRM_MOUSE			
35	SRM_MOUSE			
36	SRM_MOUSE			
37	SRM_MOUSE			
38	SRM_MOUSE			
39	SRM_MOUSE			
40	SRM_MOUSE			
41	SRM_MOUSE			
42	SRM_MOUSE			
43	SRM_MOUSE			
44	SRM_MOUSE			
45	SRM_MOUSE			

RT complete genome sequence. "

[illegible]

```

Ddb      219 VCGAHLACALHMLMNAEASGSGNLLKRVKSYVW 26
          || | | | | | | | | | | | | | | | | |
Gy       14 VCNHGSSMGHGLIYINLNGTLEMSMT-VM 49
          | 15| Conserved 70 72 Mismatches 70

RESULT   15
ID        YCUB_6_YEAST             STAMANDIC    Pri1     566 AA.
AC        P25370;
CD        01 MAY-1992 (vol. 22, state update)
DI        01 MAY-1992 (Vol. 22, last sequence update)
DE        15-DEC-1998 (Vol. 37, last annotation update)
DE        HYPOHECTICAL 64.9 KDA PROTEIN IN SKOV-3IP5 INTER
GN        YCUB36W OR YCUB16W OR YCUB187
OS        Saccharomyces cerevisiae (Baker's yeast).
OC        Eukaryota; Fungi; Ascomycota; Saccharomycetes; Sa
OR        Saccharomycetaceae; Saccharomycos.
RN        [1]
RP        SEQUENCE FROM N.A.
RX        MEDLINE: 91377317.
FA        Rat R. f. foetalis clone R. X. J. Kleinhaus et al.
RT        Full complete sequence of a 1.4 kb fragment of
KT        III encompasses four novel coding frames.
KL        Yeast 7:533-536(1991).

CC        This SWISS-Pro entry is obsolete. It is proteo
CC        between the Swiss Institut Bioinformatics at
CC        the European Bioinformatics Institute. There are
CC        use by non profit institutions as long as its
CC        modified and this statement is not removed. Users
CC        entities requires a license agreement (See: http://
CC        or send an email to lic@ebi.ac.uk for details.)
CO
EMBL: X594720; CAA4206.1;
DR        PIR: S19364; S19364.
DR        PIR: S17477; S17477.
KW        Hypothetical protein
SC        SH003870; 566 AA; 64.9 kDa; 68% identity to
Query Match      18.4%; Score 73; DR 1; Length
Best Local Similarity 42.0%; Prod. No. 2,600,000
Matches      9; Conservation 4; Mismatches 72

```

Search completed: Fri Aug 27 11:11:2009
Job time : 12 secs.

[illegible]




```

FN [2]
SEQUENCE FROM N.A.
RP STRAIN-ANT7:
KA HLOUIN J.C.:
RL Submitted (JAN 1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U84100; AAB58535.1;
DR PFAM: PF00516; GP120; 1.
DR PFAM: PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 854 AA; 97975 MW; 79577AF3 CRC32;

Query Match
Best Local Similarity 46.48; Score 83; DB 14; Length 854;
Local Similarity 46.48; Pred. No. 1,66e-01;
Matches 12; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

DB 445 SCSSSTICIMLCQIYNVSVSVAARVAQWR 467
||||| 11: |||| 11: |||| 11: |||| 11: ||||
19 GSSSAI-GILEAGQIYNSNLSMSPEWR 50

RESULT 11
ID 012006; PRELIMINARY; PRT; 866 AA.
AC 012006;
DT 01-JUL-1997 (Tremblé 04, Created)
DE 01-JUL-1997 (Tremblé 04, Last sequence update)
DE 01-NOV-1998 (Tremblé 08, Last annotation update)
DE ENVELOPE GLYCOPROTEIN GP160.
CN HIV-1.
OS Chimpazee immunodeficiency virus (SIVcpz) (CTV).
OC Viruses; Retroviridae; Retrovirales; Lentivirinae;
LN [1]
RP SEQUENCE FROM N.A.
KA NYAHM P., LEWIS P., FIELDS M., JANSSENS W., HERNIMANN J.,
KA FRANKEN K., ANDRIES K., DEN HAESEVELD M., HERNY J., PIOT P.,
KA DER GROEN G.:
RL J Virol. 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
KA STRAIN-ANT5:
KA HLOUIN J.C.:
RL Submitted (JAN 1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U84098; AAB58534.1;
DR PFAM: PF00516; GP120; 1.
DR PFAM: PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 866 AA; 94954 MW; 10CF2951 CRC32;

Query Match
Best Local Similarity 46.48; Score 83; DB 14; Length 866;
Local Similarity 46.48; Pred. No. 1,66e-01;
Matches 12; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

DB 446 SCSSSTICIMLCQIYNVSVSVAARVAQWR 478
||||| 11: |||| 11: |||| 11: |||| 11: ||||
19 GSSSAI-GILEAGQIYNSNLSMSPEWR 50

RESULT 12
ID 012009; PRELIMINARY; PRT; 868 AA.
AC 012009;
DT 01-JUL-1997 (Tremblé 04, Created)
DE 01-JUL-1997 (Tremblé 04, Last sequence update)
DE 01-NOV-1998 (Tremblé 08, Last annotation update)
DE ENVELOPE GLYCOPROTEIN GP160.
CN HIV-1.
OS Chimpazee immunodeficiency virus (SIVcpz) (CTV).
OC Viruses; Retroviridae; Retrovirales; Lentivirinae;
LN [1]
RP SEQUENCE FROM N.A.
KA STRAIN-ANT6:
KA NYAHM P., LEWIS P., FIELDS M., JANSSENS W., HERNIMANN J.,
KA FRANKEN K., ANDRIES K., DEN HAESEVELD M., HERNY J., PIOT P.,
KA DER GROEN G.:
RL J Virol. 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
KA STRAIN-ANT5:
KA HLOUIN J.C.:
RL Submitted (JAN 1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U84098; AAB58534.1;
DR PFAM: PF00516; GP120; 1.
DR PFAM: PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 866 AA; 94954 MW; 10CF2951 CRC32;

```

```

RL J. Virol. 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
KA STRAIN-ANT8:
KA HLOUIN J.C.:
RL Submitted (JAN 1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U84101; AAB58536.1;
DR PFAM: PF00516; GP120; 1.
DR PFAM: PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 868 AA; 99508 MW; DA13BDE CRC32;

Query Match
Best Local Similarity 36.14; Score 83; DB 14; Length 868;
Local Similarity 36.14; Pred. No. 1,66e-01;
Matches 12; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

DB 449 SCSSSTICIMLCQIYNVSVSVAARVAQWR 481
||||| 11: |||| 11: |||| 11: |||| 11: ||||
19 GSSSAI-GILEAGQIYNSNLSMSPEWR 50

RESULT 13
ID 09X298; PRELIMINARY; PRT; 2081 AA.
AC 09X298;
DT 01-NOV-1999 (Tremblé 12, Created)
DE 01-NOV-1999 (Tremblé 12, Last sequence update)
DE 01-NOV-1999 (Tremblé 12, Last annotation update)
DE L402.11. PROTEIN.
CN L402.11.
OS Leishmania major.
OC Eukaryota; Eukaryota; Kinetoplastida; Trypanozomatidae; Leishmania.
LN [1]
RP SEQUENCE FROM N.A.
KA MURRAY E., GUATE M., LAWSON J., HARRIS D., FALANKEAM M., IVINS A.,
KA HARRILL H.:
RL Submitted (MAY 1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL031909; CAB42353.1;
DR PFAM: PF00517; GP41; 1.
KW Envelope protein.
SQ SEQUENCE 2081 AA; 224089 MW; F1037DEB CRC32;

Query Match
Best Local Similarity 37.19; Score 83; DB 5; Length 2081;
Local Similarity 37.19; Pred. No. 1,66e-01;
Matches 13; Conservative 10; Mismatches 11; Indels 2; Gaps 2;

DB 778 HGCGSG-SRACGGAAGAAALHVAWSS-DWRER 810
||||| 11: |||| 11: |||| 11: |||| 11: ||||
18 HGCGSAGILFACQIYNSNLSMSPEWR 52

RESULT 14
ID 09YB79; PRELIMINARY; PRT; 683 AA.
AC 09YB79;
DT 01-NOV-1999 (Tremblé 12, Created)
DE 01-NOV-1999 (Tremblé 12, Last sequence update)
DE 01-NOV-1999 (Tremblé 12, Last annotation update)
DE 683AA LONG HYPOTHETICAL PROTEIN.
CN APEL599.
OS Acetopyrum pernix.
OC Archaea; Crenarchaeota; Acetopyrum.
LN [1]
RP SEQUENCE FROM N.A.
KA STRAIN-K1:
KA MEDLINE: 99310339.
KA FAWCZARZAKI V., HINO Y., HOSIYAMA H., YAMAZAKI S., HAIKAWA Y.,
KA JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H.,
KA HOSIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H.,
KA YAMAMOTO M., MAEDA S., ENMAHASE T., IANAKA I., FUKUI Y.,
KA YAMAMOTO J., YOSHIDA N., YOSHIDA A., NAKAI Y., YOSHIDA K., KAWAMURA Y.,
KA NOMURA N., SAKO Y., KIKUCHI H.:
RL Complete genome sequence of an aerobic hyper-thermophilic
RL crenarchaeon, Acetopyrum pernix K1.
RL DNA Res. 6:83-101(1999).
DR EMBL: AF200062; BAA80599.1;
SQ SEQUENCE 683 AA; 75229 MW; 447B052 CRC32;

```

Query Match 20.6% Score 82; DB 1; Length 683;
Best Local Similarity 36.4%; Pred. No. 2,46e-01;

Matches 12; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

DB 641 SEVGVIRHGGVAVSSVLAHMAHMAHRRR 663

UY 22 SAIGLIP-AGOTVGNQSLISMSTVEMRRR 54

RESULT 15

AC 01-2005; PRELIMINARY; PK1; 866 AA.

DI 01-JUL-1997 (ITEMBREL 04, created)

DI 01-JUL-1997 (ITEMBREL 04, last sequence update)

DI 01-NOV-1998 (ITEMBREL 08, last annotation update)

DE ENVELOPE GLYCOPROTEIN GP160.

ENV.

OS Chimpanzee immunodeficiency virus (SIVcpz) (TIV).

OC Viruses; Retroviruses; Retroviridae; Lentiviruses.

KN 11

RP SEQUENCE FROM N.A.

RC STRAIN-AN14;

KA NYAMBI P., LEWIS P., PETERSEN M., JANSSENS W., HEYNDRICKX L.,

KA FRANSSEN K., ANDERSEN K., DEN HARTVELDE M., HEENEY J., PLOU P.,

KA DER GROEN G.;

RL J. VIROL. 0:0-0(0).

KN 12

RP SEQUENCE FROM N.A.

RC STRAIN-AN14;

KA BL001N J.C.;

RL Submitted (JAN-1997) to the EMBL/GenBank/Tranbase.

DR EMBL: D84097; AAB58542.1;

DR PRAM: PF00516; GP120.1;

DR PRAM: PF00517; GP41.1;

KW Envelope Protein.

SV SEQUENCE 866 AA; 99373 MW; ACF20B2D CHG32;

Query Match 20.6% Score 82; DB 14; Length 866;

Best Local Similarity 36.4%; Pred. No. 2,46e-01;

Matches 12; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

DB 445 SCSSIIIGIMEDJYVNVNSVPAVAVDQWR 477

UY 19 GGSSEAI-GLEVA3GYGNSNQLSLSMSDPEWR 50

Search completed: Fri Aug 25 18:09:11 2000
Job time : 40 secs.



 RELEASE 3.1A John F. Collins, BioComputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

 (TM)

Search: protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Aug 25 18:03:56 2000; Maspar time 4.74 seconds
 Tabular output not generated. 338 281 Million cell updates/sec

Title: >US-09-257-585-7
 Description: (1-34) from US09257585.pcp
 Perfect Score: 250
 Sequence: 1 MPPAIPFOLYPTIRSFALITLHQLITITMSLVQ 34

Scoring table:
 Gap 11
 PAM 150
 Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: p1r64
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 34.449; Variance 63.944; Scale 0.539

Prod. No. is the number of results prioritized by e-value to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Prod. No.
1	94	37.6	645	2	T1137	NADH dehydrogenase su	1,948-02
2	92	32.8	289	2	T42596	glucose-6-phosphate	6,548-01
3	82	32.8	441	1	JX0364	antithrombin III	6,558-01
4	81	32.4	452	1	A45920	urate oxidase (EC 1.7	9,086-01
5	80	32.0	428	1	DBH057	estradiol 17beta-dehy	1,260-00
6	80	32.0	465	2	159611	antithrombin III - mo	1,260-00
7	79	31.6	422	2	G72643	probable transketolas	1,760-00
8	77	30.8	144	2	G71295	hypothetical protein	3,328-00
9	77	30.8	318	2	T11441	NADH dehydrogenase (u	3,328-00
10	77	30.8	466	2	A25330	bacteriothiochrome	3,328-00
11	77	30.8	464	1	XHH03	antithrombin III prec	4,320-00
12	76	30.4	442	1	S29133	urate oxidase (EC 1.7	4,570-00
13	75	30.0	439	2	A50077	obligated-cytochrome	6,260-00
14	74	29.5	242	2	T51542	probable ribose 5 pho	8,550-00
15	74	29.5	522	2	S60362	glucose transport pro	8,550-00
16	74	29.6	522	2	A41556	glucose transport pro	8,550-00
17	74	29.6	591	1	ACT481	nicotinyl acetylcholi	8,550-00
18	74	29.6	691	2	A44863	band 4.2 protein - hu	8,550-00
19	74	29.6	741	2	A44865	protein 4.2 human	8,550-00
20	74	29.4	282	2	E71543	hypothetical protein	1,160-01
21	73	29.4	479	1	S17408	obligated-cytochrome	1,160-01
22	73	29.4	465	1	S28219	antithrombin III prec	1,160-01
23	73	29.2	523	2	S06920	glucose transport pro	1,160-01

24	72	28.8	448	2	G76724	thioamide
25	72	28.8	443	2	G72652	estradiol
26	72	28.8	444	2	S87962	estradiol
27	72	28.8	444	2	G71485	antithrom
28	72	28.8	443	2	H71612	aspartate
29	72	28.8	443	2	H71612	aspartate
30	71	28.4	466	2	S45806	hypothet
31	71	28.4	479	2	A71499	hypothet
32	71	28.4	443	1	J05166	ketol-acid
33	71	28.4	443	2	S61143	EMO-prote
34	71	28.4	479	2	S48155	obligated
35	71	28.4	495	2	E70622	probable
36	71	28.4	601	2	T11451	NADH's prot
37	71	28.4	749	2	H97604	Two-compo
38	71	28.4	749	2	A45848	compone
39	71	28.4	901	2	I11155	hypothet
40	71	28.4	3985	2	I00347	polyprot
41	70	28.0	423	2	I15815	sigma 700
42	70	28.0	423	2	I15815	sigma 700
43	70	28.0	440	2	S42919	protein
44	70	28.0	440	2	S42919	obligated
45	70	28.0	1462	2	H01440	hypothet

REFERENCES

RESULT 1
 ENTRY T1137 #Type: complete
 TITLE NADH dehydrogenase subunit 5 - access
 ORGANISM #Formal name: mitochondrion Rattus norvegicus
 DATE 16-Jul-1999 #Sequence revision 1: 142

ACCESSIONS
 REFERENCE T1137
 #Authors Castresana, J., Belamater-Furios, J., V
 Paabo, S.
 Z17250
 T1137
 20-Sep-1999

#Journal Genetics (1999) 151:1115-1124
 #Title The mitochondrial genome of the house
 #References Castresana, J., Belamater-Furios, J., V
 Paabo, S. 1999. The evolution of double
 #Cross-References M010000000

#Accession T1137
 #Status preliminary, translated from 5c2
 #Molecule_Type DNA
 #Residues 1445, #Total 148
 #Cross-References EMBL:AB011945.1
 F010N:AA011945.1

GENETICS
 #Genome mitochondrion
 #Classification #Support: NADH dehydrogenase (ubiq
 #Keywords mitochondrion
 #Summary #Length 645 #molecular weight 69k4

Query Match
 Host local Similarity 69.7% Prod. No. 1,948-02
 Matches 17, Conserved 5, Misaligned 72

ID 45 EUCALYPTUS-TPR-PROTEIN-INDIVIDUAL 72
 1 1 11 1 141 1111 1 11 111
 97 3 FAITGGLTTRKASTATLALALITLALV 32

RESULT 2
 ENTRY D42586 #Type: fragment
 TITLE glucose-6-phosphate dehydrogenase (EC 2.4.1.17,
 (transient))
 ORGANISM #Formal name: Homo sapiens #Common name
 HUMAN
 DATE 04-Mar-1993 #Sequence revision 15: N02

ACCESSIONS
 REFERENCE D42586
 #Authors Rutter, R. K., Yoon, E., Shown, K. Y.,
 Yeaman, M. L., O'Connell, J. S.,
 J. Biol. Chem. (1992) 267:14247-14254

```

#title
A novel complex locus D31 encodes human fibrinogen, phenol
and other UDP-glucuronosyltransferase isozymes with
identical carboxyl termini.

#cross-references
MIM:92147680

#accession
D42586

#status
Preliminary; not compared with conceptual translation
#molecule_type
nucleic acid
#residues
1-289 #label R17
#cross-references
db_MIM:127; R17:34014; P1N:AAA92020.1; P1D:3440135
#note
Sequence extracted from NBTI cDNA (NBTI-R144)

CLASSIFICATION
#superfamily
glucuronosyltransferase
#glycosyltransferase
hexosyltransferase
#length
289 #checksum
4342

KEYWORDS
SUMMARY

Query Match: 42.8% Score 827 E0.2 Length 289
Best Local Similarity 46.4% Pred. No. 6,536-01;
Matches 14; Conservative 7; Mismatch 5; Indels 0; Caps 3;

DB 97 JOURNAL ELECTRONIC JN:V 127
QY 7 TOLYPTRISFLA-ILHOLI-LITMSLV 42

RESULT 3
ENTRY
#type
complete
#name
antithrombin III - pig
#formal_name
Sus scrofa domestica #common_name
domestic pig
#date
10-Sep-1999 #sequence_revision
10-Sep-1999 #text_change
10-Sep-1999
ACCESSIONS
JX0364
REFERENCE
JX0464
#authors
Tokunaga, F.; Goto, T.; Wakabayashi, S.; Koide, T.
#journal
J. Biochem. (1994) 116:1164-1170
#title
Amino acid sequence of porcine antithrombin III.
#cross-references
MIM:95204393
#accession
JX0364
#molecule_type
protein
#residues
1-431 #label IOK

FUNCTION
#description
Inhibits in blood plasma thrombin and activated coagulation
factor X, either weakly alone or strongly in the presence
of heparin
#superfamily
antithrombin III
#keywords
glycoprotein; heparin binding
FEATURE
#region
hinge #status
predicted
#disulfide_bonds
#status
predicted
#binding_site
carboxylate (Asn) (covalent) #status
experimentally
#inhibitory_site
Arg (thrombin) #status
predicted
#length
431 #molecular_weight
48936 #checksum
462

344
SUMMARY

Query Match: 42.8% Score 827 E0.2 Length 431
Best Local Similarity 44.2% Pred. No. 6,536-01;
Matches 8; Conservative 17; Mismatch 8; Indels 0; Caps 0;

DB 94 RSLNPARVTEKANRETVLIRVALNTIFRGR 426
QY 2 RQATITQALPTTRDSFLATIHQITLITMSIVQ 44

RESULT 4
ENTRY
#type
complete
#name
urate oxidase (EC 1.7.3.3) - fruit fly (Drosophila
melanogaster)
#formal_name
Drosophila melanogaster
#date
09-Nov-1990 #sequence_revision
27-Jun-1994 #text_change
14-Aug-1999
ACCESSIONS
A45920; S08676
REFERENCE
A45920
#authors
Wallbrink, L.L.; Burnett, J.B.; Friedman, Y.B.

```

```

#journal Mol. Cell. Biol. (1990) 10:5114-5127
#title Molecular characterization of the phenophita melanogaster
        acute catalase gene, an enzyme replicable gene expressed
        only in the malpighian tubules.
cross-references MIMD:90377201
accession AF5920
#molecule-type DNA
#residues 1-352 ##label MAL
##cross-references EMBL:M1143, NID:43736, FID:GMA6283.L, FID:48799
SERIES
#sequence (0)
#cross-references FlyBase:Fbm000361
#inrons 192/1
CLASSIFICATION #superfamily male oxidase
FEATURE #copper binding, oxidoreductase, peroxisome
FE0152 #region: peroxisomal; 74, 75, same location signal
              (S-PKKH-F) motif N
              #hydrophobic copper (his) (type 2) #status predicted
SUMMARY #length 352 #molecular-weight 40034 #checksum 5711
Query March 32.4% score 61; 19.1; length 352;
Match Local Similarity 27.6%; Pred. Num. 9,08e-01;
Matches 8; Conservative 12; Mismatches 8; Indels 1; Gaps 1;

Db 277 PSYCHITVLS-SEROVINDIPCVSVISM 304
||| ||| ||| ||| ||| |||
Oy 3 PALPOLYPEPTORSFLALIHQITITMSL 31

RESULT 5
ENTRY DEHQ67 #type complete
TITLE estradiol 17beta-dehydrogenase (EC 1.1.1.62) type I - human
ALTERNATE_NAMES 17beta-estradiol dehydrogenase, 17beta hydroxysteroid
                dehydrogenase
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31 Mar 1992 #sequence_revision 31 Mar 1992 #text_change
ACCESSION A24081; S29286; S81654; S47212; A40146; M60440, A21031;
REFERENCE A60628; C02067, A37190, A37239
#authors Liu Jhe, V.; Labile, C.; Simard, J.; Lachance, Y.; Zhao,
        H.F.; Couet, D.; Leblanc, G.; Labile, P.
#journal Mol. Endocrinol. (1990) 4:268-275
#title Structure of two in tandem human 17beta-hydroxysteroid
        dehydrogenase genes.
cross-references MIMD:9023140
accession A56081
#molecule-type DNA
#residues 1-328 ##label LUU
##cross-references GB:A27138; NID:9181950; PIDN:AB16941.1; PID:q381951
REFERENCE S29286
#authors Pollocko, B.; Isomaa, V.; Vilko, R.
#journal Eur. J. Biochem. (1992) 205:459-466
#title Genomic organization and DNA sequences of human
        17beta-hydroxysteroid dehydrogenase genes and flanking
        regions. Localization of multiple Alu sequences and
        putative cis-acting elements.
cross-references MIMD:9401163
accession S29288
#molecule-type DNA
#residues 1-328 ##label PIL
##cross-references EMBL:M84472; NID:980692; FIDN:AMD1642.1;
        PID:q177127
REFERENCE S01654
#authors Felicketo, B.; Isomaa, V.; Marenlausta, O.; Vilko, R.
#journal FEBS Lett. (1988) 239:73-77
#title Complete amino acid sequence of human placental
        17 beta-hydroxysteroid dehydrogenase deduced from cDNA.
cross-references MIMD:89031221

```

```

#accession      S01654
##molecule_type      mRNA
##residues      2-328 ##label      PEZ2
##cross-references      CHEBI(3449) NID:q97584; EMBL(ACAA179.1), FID:q97805
#accession      S43212
##molecule_type      protein
##residues      2-24758-58;205-223 ##label      PE2
REFERENCE
#authors      AAO146
            Ioo Thor, V., Labrie, C., Zhang, H.W., Gossel, J., MacLennan, Y.,
            Simard, J., Leblanc, G., Goto, J., Horobe, D., Gagne, P.,
            Labrie, F.
#journal      Mol. Endocrinol. (1989) 3:1301-1309
#title      Characterization of cDNAs for human estradiol
            17beta-dehydrogenase and assignment of the gene to
            chromosome 17: evidence of two mRNA species with distinct
            5'-terminal in human placenta.
##cross-references      M01D:8914467
#accession      AAO146
##molecule_type      mRNA
##residues      1-328 ##label      THE
##cross-references      CR:M77148 NID:q181450; PIN:AA16441 1; PIN:q181951
REFERENCE
#authors      AB0440
            Gast, M.J., Sims, H.E., Murdoch, G.L., Gast, P.M., Strauss,
            A.W.
#journal      Am. J. Obstet. Gynecol. (1989) 161:1726-173.
#title      Isolation and sequencing of a complementary deoxyribonucleic
            acid clone encoding human placental 17beta-estradiol
            dehydrogenase: identification of the putative cofactor
            binding site.
##cross-references      M01P:00102621
#accession      AB0440
##status      not compared with conceptual translation
##molecule_type      mRNA
##residues      5-66;67-70-94;79AAGCAAGT;102-158;160-178;177,181-311,
            319,313-328 ##label      GAS
REFERENCE
#authors      A21931
            Murdoch, G.L., Chin, C.C., Clifford, K.E., Bradshaw, K.A.,
            Warren, J.C.
#journal      J. Biol. Chem. (1983) 258:11460-11464
#title      Human placental estradiol 17-beta-dehydrogenase.
            Identification of a single histidine residue
            critically labeled by both 3-bromodeoxyestrone and
            17beta-estrone-oxy-4-estrone-5,17-dione.
##cross-references      M01D:84008145
#accession      A21931
##molecule_type      protein
##residues      229-224 ##label      M98
REFERENCE
#authors      AB0628
            Murdoch, G.L., Chin, C.C., Warren, J.C.,
            Biochemistry (1986) 25:641-646
#journal      Human placental estradiol 17beta-dehydrogenase: sequence of a
            histidine-bearing peptide in the catalytic region.
##cross-references      M01D:8615978
#accession      AB0628
##molecule_type      protein
##residues      208-215 ##label      M02
REFERENCE
#authors      A37190
            Labrie, F., Iuorio, V., Labrie, C., Horobe, D., Gossel, J.,
            Zhao, H.F., Gagne, P., Simard, J.,
            J. Steroid Biochem. (1989) 34:189-197
#journal      Characterization of two mRNA species encoding human estradiol
            17beta-dehydrogenase and assignment of the gene to
            chromosome 17.
##cross-references      M01D:90174172
#contents      annotation
#accession      G09150
#authors      Shen, Y.
#submission      submitted to the EMBL Data Library, August 1995
#accession      G02067
##status      translated from CR/MML/PMML
##molecule_type      DNA
##residues      1-328 ##label      SHE
##cross-references      EMBL:U14879; NID:q97584; PIN:AA05019.1;

```

[illegible]


```

ACCESSIONS      D71542
REFERENCE        A71570
AUTHORS          Stephens, K.S.; Kalmady, S.; Lammel, C.J.; Fero, J.; Marathe,
                  R.; Aravind, L.; Mitchell, W.P.; Ingrey, L.; Tatunov,
                  R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
JOURNAL          Science (1999) 282:754-759
TITLE            Genome sequence of an obligate intracellular pathogen of
                  humans: Chlamydia trachomatis.
CROSS-REFERENCES MIMD:9900809
ACCESSION        D71542
STATUS           preliminary
MOLECULE_TYPE    DNA
RESIDUES         1-242 ##label AKN
CROSS-REFERENCE DB:AE012945; DB:AE01273; MIM:93328617; FID:93328621
EXPERIMENTAL_SOURCE serotype D, strain UW-3/7x

GENETICS
GENE              TPIA
CLASSIFICATION    #superfamily Haemophilus influenzae ribose-5-phosphate
                  isomerase
SUMMARY           #length 242 #molecular-weight 26646 #checksum 8729

Query Match      29.8% Score 74; 198.2; length 242;
Best Local Similarity 34.88; Pred. No. 8,55e+00;
Matches          8; Conservative 11; Mismatches 3; Indels 1; Gaps 1;

Id               9 PSSIK-EEPIKLEKLEAVAVE <0
CY              11 PPIHRSFLAIHQIILIMSLVD 34

ENTRY            15
TITLE            glucose transport protein type 2 - rat (fragment)
ORGANISM         15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change
DATE             07-May-1999
ACCESSIONS       S68362
REFERENCE        S68362
AUTHORS          Ahn, Y.; Kim, J.; Han, G.; Lee, H.; Kim, Y.
JOURNAL          Arch. Biochem. Biophys. (1995) 323:387-396
TITLE            Cloning and characterization of rat pancreatic
                  beta-cell/11c-1 type glucose transporter gene: a unique
                  exon/intron organization.
CROSS-REFERENCES MIMD:96063615
ACCESSION        S68362
STATUS           preliminary; nucleic acid sequence not shown;
                  translation not shown
MOLECULE_TYPE    DNA
RESIDUES         1-522 ##label AHN
CROSS-REFERENCE EMBL:L28134
NOTE             The nucleotide sequence was submitted to the EMBL Data
                  Library, September 1994

GENETICS
EXONS            5/7 6/7 1-22 1-4 7 20 6 20 1 11 3 524 6 126 4;
                  456/3
CLASSIFICATION    #superfamily glucose transport protein
SUMMARY           #length 522 #checksum 4585

Query Match      29.6% Score 74; 198.2; length 522
Best Local Similarity 50.08; Pred. No. 8,55e+00;
Matches          12; Conservative 4; Mismatches 7; Indels 1; Gaps 1;

Id               179 PPIHSGAICTHQIAIVTIIISQ 202
CY              12 PPIHRSFLAIHQIILIMSLVD 34

Search completed: Fri Aug 25 18:04:28 2000
Job time : 32 secs.

```



CC	TOPOLGY:	linear
CC	MOLECULE TYPE:	protein
CC	SEQUENCE:	574 AA; 57862 PW; 155107 CH;
SO		
CC	Query Match	27.6%; Score 592; DB 2; Length 534;
CC	Best Local Similarity	45.48%; Pred. No. 7.12e+01;
CC	Methods	21; Conservative; 5; Mismatched 7; Indels 1; Gaps 1
IDB		
DB	193 P1A1RCAICLHQAIVGIIHSQ 216	
UY	11 11 111111 1 1 1	
UY	12 PID-RSHALHQLHTMTSLVDQ 34	
RESULT	8	
ID	US-08-461-599-17	STANDARD; PRI: 525 AA.
XX		
AC	*****	
XX		
XX		
DE	Sequence 17, Application US/08461599	
CC		
CC	Sequence 17, Application US/08461599	
CC	Patent No. 5658777	
CC	GENERAL INFORMATION:	
CC	APPLICANT: FUJII, Toshio	
CC	APPLICANT: IWAMATSU, Akihito	
CC	APPLICANT: YOSHIMOTO, Hiroyuki	
CC	APPLICANT: MINETOKI, Toshiyaka	
CC	APPLICANT: BUKAKI, Takayuki	
CC	APPLICANT: NAGASAWA, Naoshi	
CC	TITLE OF INVENTION: ALCOHOL ACETYLTANSFERASE GENES AND USE	
CC	TITLE OF INVENTION: THEROOF	
CC	NUMBER OF SEQUENCES: 19	
CC	CORRESPONDENCE ADDRESSES:	
CC	ADDRESSEE: Foley & Lardner	
CC	STREET: 3000 K Street, N.W., Suite 500	
CC	CITY: Washington, D.C.	
CC	COUNTRY: USA	
CC	ZIP: 20007-5109	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: floppy disk	
CC	COMPUTER: IBM PC compatible	
CC	OPERATING SYSTEM: PC-TOS/MS-DOS	
CC	SOFTWARE: Patent Release #1.0, Version #1.25	
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER: US/08/461,599	
CC	FILING DATE: 05-JUN-1995	
CC	CLASSIFICATION: 435	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER: US 08/077,939	
CC	FILING DATE: 18-JUN-1993	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER: JP 62497/1993	
CC	FILING DATE: 26-FEB-1993	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER: JP 184328/1992	
CC	FILING DATE: 18-JUN-1992	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: BENI, Stephen A.	
CC	REGISTRATION NUMBER: 29,768	
CC	PREFERENCE/PACKET NUMBER: 49441/195 KYPA	
CC	TELECOMMUNICATIONS INFORMATION:	
CC	TELEPHONE: (202)672-5300	
CC	TELEFAX: (202)672-5399	
CC	TELEX: 904136	
CC	INFORMATION FOR SEQ ID NO: 17:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 525 amino acids	
CC	TYPE: amino acid	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: protein	
CC	SEQUENCE: 525 AA; 61064 MW; 1459439 CN;	
SO		

RESULT 11
ID US-08-465-344-17 STANDARD: PRT: 525 AA.
AC XXXXXX
XX
XX
Sequence 17, Application US/08465334
Sequence 17, Application US/08465334
Patent No. 5728412
GENERAL INFORMATION:
APPLICANT: FUJII, Toshio
APPLICANT: IWAMATSU, Akihito
APPLICANT: YOSHIMOTO, Hiroyuki
APPLICANT: MINETOKI, Toshiyuki
APPLICANT: HOGAKI, Takayuki
APPLICANT: NAGASAWA, Naoshi
TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US-08/465,334
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US-08/077,939
FILING DATE: 18-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 62947/1993
FILING DATE: 26-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 184328/1992
FILING DATE: 18-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/101 KYPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE: 525 AA; 61064 MW; 1459439 CN;
Query Match: 26.8%; Score 67; DB 1; Length 525;
Best Local Similarity 25.0%; Pred. No. 1,02e+02;
Matches 5; Conservative 10; Mismatches 5; Indels 0; Gaps 0;
DB 106 HDTISVQELKISGVLINEQ 125
QY 15 RSPFLAIHOLLITIMSLVDQ 34
RESULT 12
ID US-08-461-621-17 STANDARD: PRT: 525 AA.

XX
AC xxxxxx
XX
XX
Sequence 17, Application US/08461621
Sequence 17, Application US/08461621
Patent No. 568284
GENERAL INFORMATION:
APPLICANT: FUJII, Toshio
APPLICANT: IWAMATSU, Akihito
APPLICANT: YOSHIMOTO, Hiroyuki
APPLICANT: MINETOKI, Toshiyuki
APPLICANT: HOGAKI, Takayuki
APPLICANT: NAGASAWA, Naoshi
TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US-08/461,621
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US-08/077,939
FILING DATE: 18-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 62947/1993
FILING DATE: 26-FEB-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 184328/1992
FILING DATE: 18-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/103 KYPA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE: 525 AA; 61064 MW; 1459439 CN;
Query Match: 26.8%; Score 67; DB 1; Length 525;
Best Local Similarity 25.0%; Pred. No. 1,02e+02;
Matches 5; Conservative 10; Mismatches 5; Indels 0; Gaps 0;
DB 106 HDTISVQELKISGVLINEQ 125
QY 15 RSPFLAIHOLLITIMSLVDQ 34
RESULT 13
ID US-08-465-344-15 STANDARD: PRT: 525 AA.
AC xxxxxx
XX
XX

DE XX Sequence 15, Application US/08465344
 DE XX Patent No. 5728412
 DE XX GENERAL INFORMATION:
 DE XX APPLICANT: FUJII, Toshio
 DE XX APPLICANT: IWAMATSU, Akihito
 DE XX APPLICANT: YOSHIMOTO, Hisayuki
 DE XX APPLICANT: MINETOKI, Yoshitaka
 DE XX APPLICANT: BOGAKI, Takayuki
 DE XX APPLICANT: NAGASAWA, Naoshi
 DE XX TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE GENES AND USE
 DE XX NUMBER OF SEQUENCES: 19
 DE XX CORRESPONDENCE ADDRESS:
 DE XX ADDRESS: Policy & Landnet
 DE XX STREET: 3000 K Street, N.W., Suite 500
 DE XX CITY: Washington, D.C.
 DE XX COUNTRY: USA
 DE XX ZIP: 20007-5109
 DE XX COMPUTER READABLE FORM:
 DE XX MEDIUM TYPE: Floppy disk
 DE XX COMPUTER: IBM PC compatible
 DE XX OPERATING SYSTEM: PC-DOS/MS-DOS
 DE XX SOFTWARE: Patent Release #1.0, Version #1.25
 DE XX CURRENT APPLICATION DATA:
 DE XX FILING DATE: 05-JUN-1995
 DE XX CLASSIFICATION: 445
 DE XX PRIORITY APPLICATION DATA:
 DE XX FILING DATE: 18-JUN-1993
 DE XX APPLICATION NUMBER: US 08/077,939
 DE XX PRIORITY APPLICATION DATA:
 DE XX FILING DATE: 26-FEB-1993
 DE XX APPLICATION NUMBER: JP 184328/1992
 DE XX ATTORNEY/AGENT INFORMATION:
 DE XX NAME: BENT, Stephen A.
 DE XX REGISTRATION NUMBER: 29,768
 DE XX REFERENCE/DOCKET NUMBER: 49441/101 KYP
 DE XX TELECOMMUNICATION INFORMATION:
 DE XX TELEPHONE: (202)672-5400
 DE XX TELEFAX: (202)672-5399
 DE XX TELETYPE: 904136
 DE XX INFORMATION FOR SEQ ID NO: 15:
 DE XX SEQUENCE CHARACTERISTICS:
 DE XX LENGTH: 525 amino acids
 DE XX TYPE: amino acid
 DE XX TOPOLOGY: linear
 DE XX MOLECULE TYPE: protein
 DE XX SEQUENCE: 525 AA; 61062 MW; 1464200 CN;
 DE XX Query Match 26.8%; Score 67; DB 1; Length 525;
 DE XX Best Local Similarity 25.0%; Prod No. 1.020e-02;
 DE XX Matches 5; Conservative 10; Mismatches 5; Indels 0; Caps 0;
 DE XX 15 RSPALHMLTITMSIVD 34
 DE XX
 DE XX RESULT 14
 DE XX ID US-08-461-621 15 STANDARD: PRI: 525 AA.
 DE XX AC xxxxxx
 DE XX XX
 DE XX Sequence 15, Application US/08461621

DE XX Sequence 15, Application US/08461621
 DE XX Patent No. 5686284
 DE XX GENERAL INFORMATION:
 DE XX APPLICANT: FUJII, Toshio
 DE XX APPLICANT: IWAMATSU, Akihito
 DE XX APPLICANT: YOSHIMOTO, Hisayuki
 DE XX APPLICANT: MINETOKI, Yoshitaka
 DE XX APPLICANT: BOGAKI, Takayuki
 DE XX APPLICANT: NAGASAWA, Naoshi
 DE XX TITLE OF INVENTION: ALCOHOL ACETYLTRANSFERASE
 DE XX NUMBER OF SEQUENCES: 19
 DE XX CORRESPONDENCE ADDRESS:
 DE XX ADDRESS: Policy & Landnet
 DE XX STREET: 3000 K Street, N.W., Suite 500
 DE XX CITY: Washington, D.C.
 DE XX COUNTRY: USA
 DE XX ZIP: 20007-5109
 DE XX COMPUTER READABLE FORM:
 DE XX MEDIUM TYPE: Floppy disk
 DE XX COMPUTER: IBM PC compatible
 DE XX OPERATING SYSTEM: PC-DOS/MS-DOS
 DE XX SOFTWARE: Patent Release #1.0, Version #1.25
 DE XX CURRENT APPLICATION DATA:
 DE XX FILING DATE: 05-JUN-1995
 DE XX CLASSIFICATION: 445
 DE XX PRIORITY APPLICATION DATA:
 DE XX FILING DATE: 18-JUN-1993
 DE XX APPLICATION NUMBER: US 08/077,939
 DE XX PRIORITY APPLICATION DATA:
 DE XX FILING DATE: 26-FEB-1993
 DE XX APPLICATION NUMBER: JP 184328/1992
 DE XX ATTORNEY/AGENT INFORMATION:
 DE XX NAME: BENT, Stephen A.
 DE XX REGISTRATION NUMBER: 29,768
 DE XX REFERENCE/DOCKET NUMBER: 49441/101 KYP
 DE XX TELECOMMUNICATION INFORMATION:
 DE XX TELEPHONE: (202)672-5400
 DE XX TELEFAX: (202)672-5399
 DE XX TELETYPE: 904136
 DE XX INFORMATION FOR SEQ ID NO: 15:
 DE XX SEQUENCE CHARACTERISTICS:
 DE XX LENGTH: 525 amino acids
 DE XX TYPE: amino acid
 DE XX TOPOLOGY: linear
 DE XX MOLECULE TYPE: protein
 DE XX SEQUENCE: 525 AA; 61062 MW; 1464200 CN;
 DE XX Query Match 26.8%; Score 67; DB 1; Length 525;
 DE XX Best Local Similarity 25.0%; Prod No. 1.020e-02;
 DE XX Matches 5; Conservative 10; Mismatches 5;
 DE XX 15 RSPALHMLTITMSIVD 34
 DE XX
 DE XX RESULT 15
 DE XX ID US-08-077-939-17 STANDARD: PRI: 525 AA.
 DE XX AC xxxxxx
 DE XX DT
 DE XX XX
 DE XX Sequence 17, Application US/0807939
 DE XX DE
 DE XX Sequence 17, Application US/0807939
 DE XX Patent No. 5521088

ID R42914 standard; Protein: 464 AA.
AC R42914;
DI 13-MAY-1994 (first entry)
DE Human antithrombin III mutant 19K.
KW AT III; mutant; serpin; protease inhibitor; serpin; thrombosis;
KW blood coagulation; anticoagulant; site-directed mutagenesis;
KW heparin binding site.
OS Homo sapiens.
FH Key location/qualifiers
FI peptide 1..32
FI label_signal_peptide
FI protein 33..464
FI /note="mature wild-type AT III"
FI region 416..430
FI /note="amino acids 384-398 of mature protein;
wild-type Val389 is substituted by Pro"
FN EP-568833-A.
PD 10-NOV-1993.
PE 08-APR-1993; 105829.
PR 10-APR-1992; JP-090488.
PR 22-FEB-1993; JP-031855.
FA (EISA) EISA CO LTD.
PI Kato H, Mizui Y, Nakajima N, Saito T, Suzuki N, Suzuki S;
Yoshihake S;
WP1: 93-352985/45.
PT New human antithrombin III mutants of high antithrombin activity
PI in absence of heparin - useful as anticoagulant for treatment
PI thrombotic disease
PS Example 3; Page 66-69; 137pp; English.
CC The invention covers mutant versions of wild-type human AT III
CC (R42995) which have at least 1 amino acid mutation in the 11-14,
CC 41-47, 125-133 and/or 384-398 regions of the mature protein. The
CC mutants have high antithrombin activity, even in the absence of
CC heparin and are useful as anticoagulants. Mutation of the heparin
CC binding site prevents interaction of the mutant AT III with
CC vascular endothelial cells; the mutants have a long in vivo
CC half-life and are resistant to inactivation by neutrophil elastase.
SQ Sequence 464 AA;
Query Match 40.8%; Score 77; DB 1; Length 464
Best local Similarity 24.2%; Pred. No. 3,070,01;
Matches 8; Conservative 15; Mismatches 10; Gaps 0;
ID 425 RSLNPNVTEKANKPELVFEVPLNTIFMGR 457
DI 13-MAY-1994 (first entry)
DE Human antithrombin III mutant 14R.
KW AT III; mutant; serpin; protease inhibitor; serpin; thrombosis;
KW blood coagulation; anticoagulant; site-directed mutagenesis;
KW heparin binding site.
OS Homo sapiens.
FH Key location/qualifiers
FI peptide 1..32
FI label_signal_peptide
FI protein 33..464
FI /note="mature wild-type AT III"
FI region 416..430
FI /note="amino acids 384-398 of mature protein;
wild-type Val389 is substituted by Ala-Ile"
FN EP-568833-A.
PD 10-NOV-1993.
PE 08-APR-1993; 105829.
PR 10-APR-1992; JP-090488.
PR 22-FEB-1993; JP-031855.
FA (EISA) EISA CO LTD.
PI Kato H, Mizui Y, Nakajima N, Saito T, Suzuki N, Suzuki S;
Yoshihake S;
WP1: 93-352985/45.
PT New human antithrombin III mutants of high antithrombin
PI activity in absence of heparin - useful as anticoagulant for
PI thrombotic disease
PS Example 5; Page 93-95; 137pp; English.
CC The invention covers mutant versions of wild-type h
CC (R42995) which have at least 1 amino acid mutation
CC 41-47, 125-133 and/or 384-398 regions of the mature
CC mutants have high antithrombin activity, even in the
CC heparin and are useful as anticoagulants. Mutatio
CC binding site prevents interaction of the mutant AT
CC vascular endothelial cells; the mutants have a long
CC half-life and are resistant to inactivation by neu
SQ Sequence 464 AA;

PI Yoshihake S;
DE WP1: 93-352985/45.
DI 13-MAY-1994 (first entry)
DE Human antithrombin III mutant 76R.
KW AT III; mutant; serpin; protease inhibitor; serpin;
KW blood coagulation; anticoagulant; site-directed mu
KW heparin binding site
OS Homo sapiens.
FH Key location/qualifiers
FI peptide 1..32
FI label_signal_peptide
FI protein 33..464
FI /note="mature wild-type AT III"
FI region 416..430
FI /note="amino acids 384-398 of mature
wild-type Val392 are substituted
by Pro-Ile"
FN EP-568833-A.
PD 10-NOV-1993.
PE 08-APR-1993; 105829.
PR 10-APR-1992; JP-090488.
PR 22-FEB-1993; JP-031855.
FA (EISA) EISA CO LTD.
PI Kato H, Mizui Y, Nakajima N, Saito T, Suzuki N,
Yoshihake S;
WP1: 93-352985/45.
PT New human antithrombin III mutants of high antithr
PI in absence of heparin - useful as anticoagulant fo
PI thrombotic disease
PS Example 5; Page 93-95; 137pp; English.
CC The invention covers mutant versions of wild type h
CC (R42995) which have at least 1 amino acid mutation
CC 41-47, 125-133 and/or 384-398 regions of the mature
CC mutants have high antithrombin activity, even in the
CC heparin and are useful as anticoagulants. Mutatio
CC binding site prevents interaction of the mutant AT
CC vascular endothelial cells; the mutants have a long
CC half-life and are resistant to inactivation by neu
SQ Sequence 464 AA;
Query Match 40.8%; Score 77; DB 1; Length
Best local Similarity 24.2%; Pred. No. 3,070,01;
Matches 8; Conservative 15; Mismatches 10; Gaps 0;
ID 425 RSLNPNVTEKANKPELVFEVPLNTIFMGR 457
DI 13-MAY-1994 (first entry)
DE Human antithrombin III mutant 76R.
KW AT III; mutant; serpin; protease inhibitor; serpin;
KW blood coagulation; anticoagulant; site-directed mu
KW heparin binding site
OS Homo sapiens.
FH Key location/qualifiers
FI peptide 1..32
FI label_signal_peptide
FI protein 33..464
FI /note="mature wild-type AT III"
FI region 416..430
FI /note="amino acids 384-398 of mature
wild-type Val392 are substituted
by Pro-Ile"
FN EP-568833-A.
PD 10-NOV-1993.
PE 08-APR-1993; 105829.
PR 10-APR-1992; JP-090488.
PR 22-FEB-1993; JP-031855.
FA (EISA) EISA CO LTD.
PI Kato H, Mizui Y, Nakajima N, Saito T, Suzuki N,
Yoshihake S;
WP1: 93-352985/45.
PT New human antithrombin III mutants of high antithr
PI in absence of heparin - useful as anticoagulant fo
PI thrombotic disease
PS Example 5; Page 93-95; 137pp; English.
CC The invention covers mutant versions of wild type h
CC (R42995) which have at least 1 amino acid mutation
CC 41-47, 125-133 and/or 384-398 regions of the mature
CC mutants have high antithrombin activity, even in the
CC heparin and are useful as anticoagulants. Mutatio
CC binding site prevents interaction of the mutant AT
CC vascular endothelial cells; the mutants have a long
CC half-life and are resistant to inactivation by neu
SQ Sequence 464 AA;

```

DB 425 RSLNPNVTEFRANKRPFLVFEVPLNTITFMGR 457
PT New human antithrombin III mutants of high antithrombin activity
in absence of heparin - useful as anticoagulant for treating
thrombotic disease
PS Example 3; Page 55-56; 137pp; English.
CS The invention covers mutant variants of wild type human AT III
(R42895) which have at least 1 amino acid mutation in the 11-14,
41-47, 125-133 and/or 384-398 regions of the mature protein. The
mutants have high antithrombin activity, even in the absence of
heparin and are useful as anticoagulants. Mutation of the heparin
binding site prevents interaction of the mutant AT III with
vascular endothelial cells. The mutants have a long in vivo
half-life and are resistant to inactivation by neutrophil elastase.
SQ Sequence 464 AA;

Query Match 30.8%; Score 77; DB 1; Length 464;
Host Local Similarity 24.2%; Pred. No. 3,07e-01;
Matches 8; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

DB 425 RSLNPNVTEFRANKRPFLVFEVPLNTITFMGR 457
PT New human antithrombin III mutants - with altered
heparin-dependent effect to factors IIA and Xa
PS Claim 2; Fig 2; 24pp; English.
DE Antithrombin 3; P-region variants; heparin-dependent; Factor IIA;
KW antithrombin 3; P-region variants; heparin-dependent; Factor IIA;
KW Factor Xa; meizothrombin; blood coagulation.
FM Factor Xa; meizothrombin; blood coagulation.
FT Key Location/Qualifiers
FT region 415..436
FT /label P-region
FT /note "Substitutions occur in this region which is
positions 383-404 in mature ATIII"

DB 425 RSLNPNVTEFRANKRPFLVFEVPLNTITFMGR 457
PT New human antithrombin III mutants of high antithrombin activity
in absence of heparin - useful as anticoagulant for treating
thrombotic disease
PS Example 3; Page 55-56; 137pp; English.
CS The invention covers mutant variants of wild type human AT III
(R42895) which have at least 1 amino acid mutation in the 11-14,
41-47, 125-133 and/or 384-398 regions of the mature protein. The
mutants have high antithrombin activity, even in the absence of
heparin and are useful as anticoagulants. Mutation of the heparin
binding site prevents interaction of the mutant AT III with
vascular endothelial cells. The mutants have a long in vivo
half-life and are resistant to inactivation by neutrophil elastase.
SQ Sequence 464 AA;

Query Match 30.8%; Score 77; DB 1; Length 464;
Host Local Similarity 24.2%; Pred. No. 3,07e-01;
Matches 8; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

DB 425 RSLNPNVTEFRANKRPFLVFEVPLNTITFMGR 457
PT New human antithrombin III mutants - with altered
heparin-dependent effect to factors IIA and Xa
PS Claim 2; Fig 2; 24pp; English.
DE Antithrombin 3; P-region variants; heparin-dependent; Factor IIA;
KW antithrombin 3; P-region variants; heparin-dependent; Factor IIA;
KW Factor Xa; meizothrombin; blood coagulation.
FM Factor Xa; meizothrombin; blood coagulation.
FT Key Location/Qualifiers
FT region 415..436
FT /label P-region
FT /note "Substitutions occur in this region which is
positions 383-404 in mature ATIII"

```

```

DB 425 RSLNPNVTEFRANKRPFLVFEVPLNTITFMGR 457
PT New human antithrombin III mutants of high antithrombin activity
in absence of heparin - useful as anticoagulant for treating
thrombotic disease
PS Example 3; Page 55-56; 137pp; English.
CS The invention covers mutant variants of wild type human AT III
(R42895) which have at least 1 amino acid mutation in the 11-14,
41-47, 125-133 and/or 384-398 regions of the mature protein. The
mutants have high antithrombin activity, even in the absence of
heparin and are useful as anticoagulants. Mutation of the heparin
binding site prevents interaction of the mutant AT III with
vascular endothelial cells. The mutants have a long in vivo
half-life and are resistant to inactivation by neutrophil elastase.
SQ Sequence 464 AA;

Query Match 30.8%; Score 77; DB 1; Length 464;
Host Local Similarity 24.2%; Pred. No. 3,07e-01;
Matches 8; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

DB 425 RSLNPNVTEFRANKRPFLVFEVPLNTITFMGR 457
PT New human antithrombin III mutants of high antithrombin activity
in absence of heparin - useful as anticoagulant for treating
thrombotic disease
PS Example 3; Page 55-56; 137pp; English.
CS The invention covers mutant variants of wild type human AT III
(R42895) which have at least 1 amino acid mutation in the 11-14,
41-47, 125-133 and/or 384-398 regions of the mature protein. The
mutants have high antithrombin activity, even in the absence of
heparin and are useful as anticoagulants. Mutation of the heparin
binding site prevents interaction of the mutant AT III with
vascular endothelial cells. The mutants have a long in vivo
half-life and are resistant to inactivation by neutrophil elastase.
SQ Sequence 464 AA;

Query Match 30.8%; Score 77; DB 1; Length 464;
Host Local Similarity 24.2%; Pred. No. 3,07e-01;
Matches 8; Conservative 15; Mismatches 10; Indels 0; Gaps 0;

DB 425 RSLNPNVTEFRANKRPFLVFEVPLNTITFMGR 457
PT New human antithrombin III mutants - with altered
heparin-dependent effect to factors IIA and Xa
PS Claim 2; Fig 2; 24pp; English.
DE Antithrombin 3; P-region variants; heparin-dependent; Factor IIA;
KW antithrombin 3; P-region variants; heparin-dependent; Factor IIA;
KW Factor Xa; meizothrombin; blood coagulation.
FM Factor Xa; meizothrombin; blood coagulation.
FT Key Location/Qualifiers
FT region 415..436
FT /label P-region
FT /note "Substitutions occur in this region which is
positions 383-404 in mature ATIII"

```

Search completed: Fri Aug 25 18:03:38 2000
Job time : 18 secs.

FI STRAND 204 210
 RA STRAND 218 229
 FI HELIX 214 237
 RT TURN 248 242
 FI HELIX 245 247
 FI STRAND 252 250
 FI STRAND 265 273
 FI TURN 274 275
 FI STRAND 277 281
 FI TURN 282 283
 FI STRAND 284 287
 FI HELIX 292 298
 FI TURN 299 303
 FI STRAND 308 319
 FI TURN 319 320
 FI STRAND 322 327
 FI STRAND 332 334
 FI TURN 335 336
 FI STRAND 337 341
 FI HELIX 343 353
 FI TURN 354 355
 FI STRAND 361 365
 FI STRAND 366 367
 SEQUENCE 466 AA: 40276 MW: 57242744E54615E1 CEE64,
 Query Match: 30.8% Score 77.48 1: Length 365
 Host Local Similarity 36.1% Pred. No. 1.24E+002
 Matches 12: Conservation 4: Mismatches 17: Indels 1: Gaps 1:
 1b 129 VERNIPITVWG-CROPHITIMVPELDNDLV 160
 1 MRPAIPOLVPIIDKSPALIIHIIITMSIWD 43
 07
 RESULT 7
 ID ANT3.HUMAN STANDARD: PRT: 464 AA.
 AC P01008:
 DT 21-JUL-1986 (Ref. 01, Created)
 DT 21-JUL-1986 (Ref. 01, Last sequence update)
 DT 15-DEC-1998 (Ref. 17, Last annotation update)
 DE ANTIHROMBIN-III PRECURSOR (AT111).
 GN AT3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 83144280.
 RA Bock S.C., Wion K.L., Vohar G.A., Lown K.M.:
 K1 "Cloning and expression of the cDNA for human antithrombin III."
 RL Nucleic Acids Res. 10:8113-8125(1982).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 83169777.
 RA Chandra L., Stackhouse K., Kidd V.J., Woo S.L.C.:
 K1 "Isolation and sequence characterization of a cDNA clone of human
 K1 antithrombin III."
 RL Proc. Natl. Acad. Sci. U.S.A. 80:1845-1848(1983).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93297227.
 RA Oids R.J., Lane D.A., Chowdhury V., de Stefano V., Leone G.,
 RA Tchin S.L.:
 K1 "Complete nucleotide sequence of the antithrombin gene: evidence for
 RT homologous recombination causing thrombophilia."
 RL Biochemistry 32:4216-4224(1993).
 RN [4]
 RP SEQUENCE OF 42-464 FROM N.A.
 RX MEDLINE: 8328456.
 RA Prochowatz E.V., Markham A.F., Orkin S.H.:
 K1 "Isolation of a cDNA clone for human antithrombin III."
 RL J. Biol. Chem. 258:8389-8394(1983).
 RN [5]
 RP SEQUENCE OF 44-464, CARBOHYDRATE-BINDING SITES, AND DISULFIDE BONDS.

RA Peterson L.F., Duda K.W., Johnson-Ka D., Sottig J.,
 RA Magnusson S.:
 K1 "Primary structure of antithrombin III (heparin C)
 RT homology between alpha 1 and alpha 2 and antithrombin
 RL (II) Colten D., Wion K., Vohar G., Oids R.J.
 K1 "The physiological functions of blood coagulation
 RL pp.43-54, Elsevier, Amsterdam (1979).
 RN [6]
 RP ACTIVE SITE.
 RX MEDLINE: 81212814.
 RA Bioeck L., Danielsson A., Fother J.W., II, Jensen
 K1 "The site in human antithrombin for functional pu
 K1 by human thrombin."
 RL FEBS Lett. 126:257-26 (1977).
 RN [7]
 RP HEPARIN-BINDING SITE.
 RX MEDLINE: 8411578.
 RA Blackbourn M.N., Smith R.L., Johnson J., Shihg J.,
 K1 "The heparin binding site of antithrombin III: an
 K1 critical tryptophan in the alpha 2 domain sequence."
 RL J. Biol. Chem. 259:9393-9400(1984).
 RN [8]
 RP METAGENESIS OF ALA 414.
 RX MEDLINE: 91192143.
 RA Austin R.C., Rastboudi F., Bock S.A., Johnson M.A.,
 K1 "Site-directed mutagenesis of amino-acid 414 of hema
 RL FEBS Lett. 250:254-256(1989).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (3.00 ANGSTROMS).
 RX MEDLINE: 94373498.
 RA Carroll R.W., Stein P.F., Fournier J., Wardell M.B.:
 K1 "Biological implications of a 3.4 A structure of the
 RL structure 2:257-270(1994).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (3.42 ANGSTROMS).
 RX MEDLINE: 95384753.
 RA Schroeder H.A., de Woot R., Finkema R., Mulders J.,
 K1 Groothuis J.D., H. W. G. J.:
 RT "The intact and cleaved human antithrombin III and
 K1 serpin-proteinase interaction."
 RL Nat. Struct. Biol. 1:44-54(1994).
 RN [11]
 RP X-RAY CRYSTALLOGRAPHY (3.00 ANGSTROMS).
 RX MEDLINE: 97220238.
 RA Skinner P., Abraham J.L., Gellera J.C., Bock S.
 K1 "The 2.6 A structure of antithrombin indicates a
 K1 change at the heparin binding site."
 RL J. Mol. Biol. 266:601-609(1997).
 RN [12]
 RP X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS).
 RX MEDLINE: 9847366.
 RA Skinner R., Chang W.S.W., de Loo P., Fei X.Y., Hsu H.,
 K1 Abraham J., Carroll R.W., Jones D.A.:
 RT "Implications for function and therapy of a 2.75 A
 K1 binary-complexed antithrombin."
 RL J. Mol. Biol. 283:9-14(1998).
 RN [13]
 RP REVIEW.
 RX MEDLINE: 91129302.
 RA Moorey J., Samama J.P., Pelatier M., Choy J.:
 K1 "Antithrombin III: structural and functional aspects"
 RL Biochimie 72:599-608(1990).
 RN [14]
 RP REVIEW ON VARIANTS.
 RX MEDLINE: 94054129.
 RA Lane D.A., Oids R.J., Fother J.W., Chowdhury V.,
 K1 Cooper D.N., Blajchman M., Fei J., Emmerich J.:
 RT "Antithrombin III mutation database: first update,
 K1 and its inhibitors Subcommittee of the Serpin Site
 K1 Committee of the International Society on Thrombosis
 K1 Haemostasis."

41 Thromb. Haemost. 70:361-369(1993).
 42 REVIEW ON VARIANTS.
 43 MEDLINE: 95269065.
 44 Stein P.E., Carrell R.W.:
 45 "What do dysfunctional serpins tell us about molecular mobility and
 46 disease?".
 47 Nat. Struct. Biol. 2:96-113(1995).
 48 REVIEW ON VARIANTS.
 49 MEDLINE: 96263733.
 50 Perry D.J., Carrell R.W.:
 51 "Molecular genetics of human antithrombin deficiency".
 52 Hum. Mutat. 7:7-22(1996).
 53 VARIANT UTAH.
 54 MEDLINE: 89050967.
 55 Hock S.C., Marrian J.A., Kadzlojowska P.:
 56 "Antithrombin III Utah: proline 407 to leucine mutation in a highly
 57 conserved region near the inhibitor reactive site".
 58 Biochemistry 27:6171-6178(1988).
 59 VARIANT TOYAMA.
 60 MEDLINE: 84119472.
 61 Koide T., Odani S., Takahashi K., Ono T., Sakuragawa N.:
 62 "Antithrombin III Toyama: replacement of arginine-47 by cysteine in
 63 hereditary abnormal antithrombin III that lacks heparin-binding
 64 ability".
 65 Proc. Natl. Acad. Sci. U.S.A. 81:289-294(1984).
 66 VARIANT BASEL.
 67 MEDLINE: 86111754.
 68 Chang J.Y., Tsai J.H.:
 69 "Antithrombin III Basel. Identification of a Pro-Leu substitution in
 70 a hereditary abnormal antithrombin with impaired heparin cofactor
 71 activity".
 72 J. Biol. Chem. 261:1174-1176(1986).
 73 VARIANT DENVER.
 74 MEDLINE: 87109210.
 75 Stephens A.W., Thalley B.S., Hiss G.H.W.:
 76 "Antithrombin-III Denver, a reactive site variant".
 77 J. Biol. Chem. 262:1044-1048(1987).
 78 VARIANT HAMILTON.
 79 MEDLINE: 89027076.
 80 Devaux K.L., Chai D.H.K., Prochownik E.V., Carter C.J.:
 81 "Antithrombin-III Hamilton: a gene with a point mutation (guanine to
 82 adenine) in codon 382 causing impaired serine protease reactivity".
 83 Blood 72:1518-1523(1988).
 84 VARIANTS GLASGOW AND NORTHWICK PARK.
 85 MEDLINE: 88189859.
 86 Erdemunt H., Lano D.A., Panico M., di Marzo V., Morris H.R.:
 87 "Single amino acid substitutions in the reactive site of antithrombin
 88 leading to thrombosis: congenital substitution of arginine 393 to
 89 cysteine in antithrombin Northwick Park and to histidine in
 90 antithrombin Glasgow".
 91 J. Biol. Chem. 263:5589-5593(1988).
 92 VARIANTS CHICAGO.
 93 MEDLINE: 8938698.
 94 Erdemunt H., Lano D.A., Panico M., di Marzo V., Morris H.R.:
 95 "Antithrombin Chicago, amino acid substitution of arginine 393 to
 96 histidine".
 97 Thromb. Res. 54:613-619(1989).
 98 VARIANT KODEN-4.
 99 MEDLINE: 90306344.
 100 Koda J.Y., Brennan S.O., Carrell R.W., George P., Perry D.J., Shaw J.:
 101 "Antithrombin Koden-IV 24 Arg-->Gys. The amino-terminal contribution

102 to heparin binding".
 103 FEBS Lett. 266:163-166(1990).
 104 VARIANT TIBETIN.
 105 MEDLINE: 91032211.
 106 Dany M., Bruck D., Perry D.J., Price J., Harper P.L., O'Meara A.:
 107 "Antithrombin Tibetan (-3 Val-->Glu): an N-terminal variant which has
 108 an aberrant signal peptidase cleavage site".
 109 FEBS Lett. 273:87-90(1990).
 110 VARIANT GENEVA.
 111 MEDLINE: 91035423.
 112 Gandrille S., Atgheb M., Lano D.A., Vissard D., Molloy-Sabatier P.:
 113 "Antithrombin Geneva, a variant with a proline 407 to leucine mutation
 114 in the reactive site".
 115 FEBS Lett. 273:87-90(1990).
 116 Note: remainder of annotations omitted.
 117 Query Match 30.88; Score 77; Pos 1; Length 464;
 118 Best Local Similarity 24.28; Ident. 100.00;
 119 Matches 4; Conservative 15; Mismatches 10; Indels 0; Gaps 0;
 120 ID 425 PSANPNVTEFANPELVTEVPLNTITMCR 457
 121 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
 122 2 KPAITGVITTSFIAIMQITIKSMVLS 54
 123 RESULT 8
 124 ID MCM2.DROME STANDARD; PRT: 887 AA.
 125 AP P49735;
 126 DT 01-OCT-1996 (Rel. 14, Created)
 127 DT 01-OCT-1996 (Rel. 14, Last sequence update)
 128 DT 15-DEC-1998 (Rel. 47, Last annotation update)
 129 DE DNA REPLICATION LICENSING FACTOR MCM2.
 130 MCM2.
 131 Prosophila melanogaster (Fruit fly).
 132 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 133 Diptera; Neoptera; Ephemeroptera; Diptera; Brachycera; Muscomorpha;
 134 Ephydroidea; Drosophilidae; Drosophila.
 135 [1]
 136 SEQUENCE FROM N.A.
 137 RE TISSUE: IVT ANTENNA; DISK;
 138 KX MEDLINE: 95347580.
 139 KA Treisman J.E., Follette P.J., O'Carroll P.H., Rodin G.M.:
 140 "Cell proliferation and DNA replication defects in a Drosophila MCM2
 141 mutant".
 142 Genes Dev. 9:1759-1715(1995).
 143 RE FUNCTION: ACT AS A FACTOR THAT LICENSE THE DNA FOR G1 AND G1/S
 144 G1. RUND OF REPLICATION PER CELL CYCLE. REQUIRED FOR DNA
 145 REPLICATION AND CELL PROLIFERATION.
 146 SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 147 -1- SIMILARITY: BELONGS TO THE MCM FAMILY.
 148 -----
 149 This SwissProt entry is copyright. It is produced through a collaboration
 150 between the Swiss Institute of Bioinformatics and the EMBL collaboration
 151 the European Bioinformatics Institute. There are no restrictions on its
 152 use by non-profit institutions as long as its content is in no way
 153 modified and this statement is not removed. Usage by and for commercial
 154 entities requires a license agreement (See <http://www.ebi.ac.uk/sequence/>
 155 or send an email to license@isb.slb.ch).
 156 -----
 157 EMBL: LA2762; AA05617.1;
 158 EMBLBASE: FR00014601; MCM2.
 159 PFM: PF00493; MCM_1.
 160 PROSITE: PS00847; MCM_1;
 161 PROSITE: PS50511; MCM_2;
 162 TRANSITION REGULATION: DNA-binding; Nuclear protein;
 163 DNA replication; Zinc-finger; ATP binding; Cell cycle.
 164 ZN.FING 314 340 MCM.
 165 DOMAIN 458 665 MCM.
 166 NP_BIND 508 515 ATP (POTENTIAL).
 167 TT SPONDGE 887 AA; 100414 MW; 26A7092194F99C8E CAC64;
 168



HA VIRGIN H.W. IV. LATERITE P.; WAMSLY P.; HALLSWORTH R.; WICK K.E.;
 HA DAL CANO A.J.; SPECK S.H.;
 KI *Complete sequence and genomic analysis of murine gammaherpesvirus
 KI 68.*
 KI J. Virol. 71:5894-5904(1997).

KN [2]
 KN SOURCE: FROM N.A.
 KN STRAIN-WDMS:
 KA LATREILLE P.; WAMSLY P.; WATERSTON R.H.;
 KL Submitted (AFR-1997) to the EMBL/GenBank/DBJ databases.
 KL EMBL: 097554; AAB6459.1;
 SO SOURCE: 1275 AA; 141954 MW; 5193780 CMC32;

Query Match 40.48; Score 76; DB 14; Length 1275;

Best Local Similarity 46.08; Pred. No. 5.25e+00;

Matches 9; Conservation 8; Mismatches 8; Indels 0; Gaps 0;

DB 310 TRPSSNLTTPPVNPNVNTSKIS 334

QY 1 MRPAIPQLYPPIDKSLATHHLL 25

RESULT 15

DB 074390 PRELIMINARY; PRT: 2609 AA.

AC 074390;

DI 01-NOV-1998 (ITEMBL: 08, Created)

DI 01-NOV-1998 (ITEMBL: 08, Last sequence update)

DI 01-NOV-1999 (ITEMBL: 12, Last annotation update)

DI POTATIVE TRANSPORT PROTEIN.

GN SPH487.16 OR SPH4286.12.06C.

OS Schizosaccharomyces pombe (Fission yeast).

OT Eukaryota; Fungi; Ascomycota; Arhizosmycetes;

OT Schizosaccharomycetales; Schizosaccharomycetaceae;

OT Schizosaccharomyces;

OT [1]

OT SOURCE: FROM N.A.

OT STRAIN-972H-7

KA LYNE M.; KALANDREAN M.A.; BARRELL H.G.; JIMENEZ MARTINEZ J;

KL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

KN [2]

KN SEQUENCE OF 2403-2609 FROM N.A.

KN STRAIN-972H-7

KA WOOD V.; KALANDREAN M.A.; BARRELL H.G.; SKELTON J.; CHURCHER C.M.;

KL Submitted (Sep 1998) to the EMBL/GenBank/DBJ databases.

IR EMBL: AL041261; CAA20412.1;

IR EMBL: AL041517; CAA20654.1;

IR PRAM: PR00400; WD40; 2;

IR PRINTS: PR00320; GPROTEINBP1.

SO SOURCE: 2609 AA; 295654 MW; 88204170 CMC32;

Query Match 40.48; Score 76; DB 3; Length 2609;

Best Local Similarity 42.98; Pred. No. 5.25e+00;

Matches 9; Conservation 5; Mismatches 6; Indels 1; Gaps 1;

DB 1837 LR-DVGLHPFKGRKSLHL 1856

QY 1 MRPAIPQLYPPIDKSLATHHLL 21

Search completed: Fri Aug 25 18:05:14 2000
 Job time : 5.2 secs.



M E S S A G E

(1*)

Release 3.1A John F. Collins, BioComputing Research Unit,
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Match_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Fri Aug 25 18:01:46 2000; Match time 7.31 seconds
Tabular output not generated. 332,083 Million cell updates/sec

Title: >08-04-257-585-6
Description: (1-35) from US09257585.pep
Perfect Score: 241
Sequence: 1 MDASINILLOOVRIIGYLVVSSIRHGTALIAA 35

Scoring table:
PAM 150
Gap 11

Searched: 225878 seqs, 69344122 residues
Post-processing: Minimum Match 04
Listing first 45 summaries

Database: splenb112
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-phc 8:sp-protocell
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-viridicoccus 14:sp-virus

Statistics: Mean 34.187; Variance 68.711; scale 0.498
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	92	37.7	541	10	041301	BETA-KETOACYL-CoA SYNTH	6.09e-02
2	87	35.7	465	2	067090	PHOSPHOTIDYLGLYCEROPH	4.09e-01
3	86	35.2	208	2	086889	CAP34EF PROTEIN.	5.63e-01
4	85	35.2	249	5	077489	HP007250 PROTEIN.	5.63e-01
5	85	34.8	185	8	055531	CYTCHROME OXIDASE SUB	7.72e-01
6	85	34.8	185	10	081850	POTATIVE PE(II) TRANSP	7.72e-01
7	85	34.8	418	2	056913	TRSA.	7.72e-01
8	84	34.0	541	11	035120	MCPLP.	1.45e+00
9	84	34.0	541	11	035120	N-ACETYLGLUCOSAMYL TRA	1.45e+00
10	84	34.0	541	11	035120	K-CHANNEL.	1.45e+00
11	84	33.6	98	8	098110	NADH DEHYDROGENASE SUB	1.97e-00
12	82	33.6	245	2	052260	PLASMIN PEPTIDOM SUBFACE	1.97e-00
13	81	33.2	486	10	064747	WOLKESS-1-LIKE PROTEIN	2.68e-00
14	81	33.2	469	1	030094	CONSERVED HYPOCHETICAL	2.68e-00
15	81	33.2	534	2	045934	STARFISH LEPIDACTES ACETIN	2.68e-00
16	81	33.2	511	2	094618	POTATIVE ALPHA-AMYLASE	2.68e-00
17	79	32.4	504	4	099524	TRANSFERRIN PROTEIN.	4.94e-00
18	78	32.4	684	2	051415	ALGinate lyase.	4.94e-00
19	78	32.0	74	2	056009	NADH DEHYDROGENASE SUB	6.66e-00
20	78	32.0	174	8	095961	CYTCHROME OXIDASE SUB	6.66e-00

21	78	32.0	174	8	050629	CYTCHROME
22	78	32.0	190	8	054555	CYTCHROME
23	78	32.0	195	8	066080	CYTCHROME
24	78	32.0	219	2	057991	SHIMIZU-1-LIKE
25	78	32.0	563	1	047977	R-ACETYLGLYCEROL
26	78	32.0	705	6	051402	TOZER-1 PRO
27	78	32.0	826	6	051604	SHIMIZU-1-LIKE
28	77	31.6	181	1	050629	ADP-RIB-SYL
29	76	31.1	177	1	050629	ADP-RIB-SYL
30	76	31.1	279	2	050629	ADP-RIB-SYL
31	76	31.1	279	2	050629	ADP-RIB-SYL
32	76	31.1	421	2	050629	ADP-RIB-SYL
33	76	31.1	864	2	050629	ADP-RIB-SYL
34	75	30.7	174	8	050629	CYTCHROME
35	75	30.7	174	8	050629	CYTCHROME
36	75	30.7	174	8	050629	CYTCHROME
37	75	30.7	174	8	050629	CYTCHROME
38	75	30.7	174	8	050629	CYTCHROME
39	75	30.7	174	8	050629	CYTCHROME
40	75	30.7	174	8	050629	CYTCHROME
41	75	30.7	174	8	050629	CYTCHROME
42	74	30.3	174	8	050629	CYTCHROME
43	74	30.3	174	8	050629	CYTCHROME
44	74	30.3	174	8	050629	CYTCHROME
45	74	30.3	174	8	050629	CYTCHROME

ACIDIMENTS

RESULT	ID	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	041301	47.78	Score 923	108	10	041301	BETA-KETOACYL-CoA SYNTH	6.09e-02
2	067090	47.78	Score 923	108	10	067090	PHOSPHOTIDYLGLYCEROPH	4.09e-01
3	086889	47.78	Score 923	108	10	086889	CAP34EF PROTEIN.	5.63e-01
4	077489	47.78	Score 923	108	10	077489	HP007250 PROTEIN.	5.63e-01
5	055531	47.78	Score 923	108	10	055531	CYTCHROME OXIDASE SUB	7.72e-01
6	081850	47.78	Score 923	108	10	081850	POTATIVE PE(II) TRANSP	7.72e-01
7	056913	47.78	Score 923	108	10	056913	TRSA.	7.72e-01
8	035120	47.78	Score 923	108	10	035120	MCPLP.	1.45e+00
9	035120	47.78	Score 923	108	10	035120	N-ACETYLGLUCOSAMYL TRA	1.45e+00
10	098110	47.78	Score 923	108	10	098110	K-CHANNEL.	1.45e+00
11	052260	47.78	Score 923	108	10	052260	PLASMIN PEPTIDOM SUBFACE	1.97e-00
12	064747	47.78	Score 923	108	10	064747	WOLKESS-1-LIKE PROTEIN	2.68e-00
13	030094	47.78	Score 923	108	10	030094	CONSERVED HYPOCHETICAL	2.68e-00
14	045934	47.78	Score 923	108	10	045934	STARFISH LEPIDACTES ACETIN	2.68e-00
15	094618	47.78	Score 923	108	10	094618	POTATIVE ALPHA-AMYLASE	2.68e-00
16	099524	47.78	Score 923	108	10	099524	TRANSFERRIN PROTEIN.	4.94e-00
17	051415	47.78	Score 923	108	10	051415	ALGinate lyase.	4.94e-00
18	056009	47.78	Score 923	108	10	056009	NADH DEHYDROGENASE SUB	6.66e-00
19	095961	47.78	Score 923	108	10	095961	CYTCHROME OXIDASE SUB	6.66e-00
20	095961	47.78	Score 923	108	10	095961	CYTCHROME OXIDASE SUB	6.66e-00

STRAIN VFE.
 EX MEDLINE: 98196666.
 FA DECKERT G., WAPPEN P.V., GASTELERLAND T., YOUNG W.C., LEMX A.L.,
 KA GRAHAM D.E., OVERBECK P., SNEAD M.A., KELLER M., ARDAY M., HIRSH P.,
 KA PRILMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 FT The complete genome of the hypothermophilic bacterium *Amplifex*
 ET *aerolius*.
 RT Nature 392:451-458(1998).
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN-VFE5.
 RA DECKERT G., WAPPEN P.V., GASTELERLAND T., YOUNG W.C., LEMX A.L.,
 KA GRAHAM D.E., OVERBECK P., SNEAD M.A., KELLER M., ARDAY M., HIRSH P.,
 KA PRILMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RL Submitted (Jun 1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF000715; AAC07048.1; -
 DR PROSITE: P500379; CDP_ALCOHOL_P_TRANSF: 1.
 DR PFAM: PF01066; CDP-CH-P_TRANSF: 1.
 SQ SEQUENCE 308 AA: 45245 MW: 4006793 CP032;

Query Match 35.7% Score 87; DB 2; Length 305;
 Best Local Similarity 34.5% Pred. No. 4,096-01;
 Matches 10; Conservative 12; Mismatches 5; Indels 2; Gaps 2;

DB 125 SAISILLIGVELGNIIIA-VIFA-VIS 151
 QY 4 ASINIIIGVVDLGYLVISITLFGHYA 31

RESULT 3
 ID 084889; PRELIMINARY; PRT; 268 AA.
 AC 084889;
 DT 01-NOV-1998 (TRENBLER, 08, Created)
 DT 01-NOV-1998 (TRENBLER, 08, Last sequence update)
 DT 01-NOV-1998 (TRENBLER, 08, Last annotation update)
 DE CAP33FE PROTEIN.
 OS CAP33FE.
 GN Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-SISP 33F TYPE 33F;
 RA LINDL D., LOPEZ R., GARCIA E., MUNOZ R.;
 RT Nucleotide sequence of the capsular gene cluster encoding the
 RT polysaccharide capsule of *Streptococcus pneumoniae* type 33F. *;
 RT Submitted (Jun-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ006986; CAA07399.1; -
 SQ SEQUENCE 268 AA: 31754 MW: 4044470 CP032;

Query Match 35.2% Score 86; DB 2; Length 268;
 Best Local Similarity 44.9% Pred. No. 5,636-01;
 Matches 13; Conservative 6; Mismatches 9; Indels 1; Gaps 1;

DB 195 LDTWEQMLITDYL-FESKIFGSELL 222
 QY 5 LNIITIGVVDLGYLVISITLFGHYA 33

RESULT 4
 ID 077489; PRELIMINARY; PRT; 269 AA.
 AC 077489;
 DT 01-NOV-1998 (TRENBLER, 08, Created)
 DT 01-NOV-1998 (TRENBLER, 08, Last sequence update)
 DT 01-NOV-1998 (TRENBLER, 12, Last annotation update)
 DE PFC0725C PROTEIN.
 OS Plasmodium falciparum.
 GN Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-3D7;
 RA MURPHY L., LAWSON D., BARRELL B.;

RL Submitted (Jun 1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z98551; CAB1145.1; -
 DR PFAM: PF01236; Form_Nit_Transf: 1.
 SQ SEQUENCE 269 AA: 29158 MW: 5416960 CP032;

Query Match 35.2% Score 86; DB 2; Length 269;
 Best Local Similarity 29.0% Pred. No. 5,636-01;
 Matches 9; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

DB 110 LAIVMALYKKVKILDY-DAVMTISLPCNYG 140
 QY 1 MDASNIIGVVDLGYLVISITLFGHYA 31

RESULT 5
 ID 035531; PRELIMINARY; PRT; 185 AA.
 AC 035531;
 DT 01-NOV-1996 (TRENBLER, 01, Created)
 DT 01-NOV-1996 (TRENBLER, 01, Last sequence update)
 DT 01-NOV-1998 (TRENBLER, 08, Last annotation update)
 DE CYTOCHROME OXIDASE SUBUNIT II (CYCMBMT).
 OS Prodoxus marthianus.
 GN Mitochondrion.
 OC Eukaryota; Metazoa; Artibeidae; Tracheata; Insecta; Insecta;
 OC Phrynosoma; Neoptera; Embryophyta; Tracheophyta; Gloriosa; Macrotrichia;
 OC Incurvarioidea; Prodoxidae; Prodoxus.
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-SANTA BARBARA CO., CA;
 RA BROWN J.K., FELLNER O., THOMPSON J.N., HARRISON R.G.;
 RT Scientific (YAH 1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U04886; AAA16080.1; -
 DR PFAM: PF00116; COX2; 1.
 KW Mitochondrion.
 LT NON-TER 185 185
 SQ SEQUENCE 185 AA: 21614 MW: 2035385 CP032;

Query Match 34.8% Score 85; DB 8; Length 185;
 Best Local Similarity 44.8% Pred. No. 7,726-01;
 Matches 13; Conservative 5; Mismatches 5; Indels 2; Gaps 2;

DB 25 DHTMILITLITVGYLL-TT-ILFNMY 51
 QY 2 DASNIIGVVDLGYLVISITLFGHYA 30

RESULT 6
 ID 081850; PRELIMINARY; PRT; 350 AA.
 AC 081850;
 DT 01-NOV-1998 (TRENBLER, 08, Created)
 DT 01-NOV-1998 (TRENBLER, 08, Last sequence update)
 DT 01-NOV-1998 (TRENBLER, 12, Last annotation update)
 DE PUTATIVE PE(II) TRANSFER PROTEIN.
 GN T1695.40.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; Eudicotyledons;
 OC Euro dicots; Rosidae; eustosids 11; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN 11
 RP SEQUENCE FROM N.A.
 RC DE IRAN M., MARSE A.C., GRIVELL L.A., BANCROFT I., MENES H.W.,
 KA MAYR K., SCHUELLER C., BRYAN M.;
 RT Submitted (Jun-1998) to the EMBL/GenBank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RC EN ARABIDOPSIS SEQUENCING PROJECT;
 RT Submitted (Jun-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A024486; CAA19685.1; -
 DR MENDEL: 32480; At4g1756; 32480.
 SQ SEQUENCE 350 AA: 37407 MW: 8942506 CP032;

Query Match 34.8% Score 85; DB 10; Length 350;

RI. New York (1997).
 RN 121
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98393797.
 RA MINDELL D.P., SORENSON M.D., DIMICHEFF D.E.:
 "Multiple independent origins of mitochondrial gene order in birds."
 RI Proc Natl Acad Sci U S A 95:10643-10647(1998)
 RN 131
 RP SEQUENCE FROM N.A.
 RA MINDELL D.P., SORENSON M.D., DIMICHEFF D.E., HASEGAWA M., AST J.C.
 YORI T.:
 "Interordinal relationships of birds and other reptiles based on whole
 mitochondrial genomes."
 RI Syst Biol 48:138-152(1999)
 RN 141
 RP SEQUENCE FROM N.A.
 RA SORENSON M.D., DIMICHEFF D.E., AST J.C., YORI T., MINDELL D.P.:
 "Complete mitochondrial DNA sequences for five birds and a turtle."
 RI Submitted (Sep-1998) to the EMBL/Genbank/Genbank databases
 RX EMBL: AF090338; AAC2500.1
 RA Mitochondrion.
 KW Mitochondrion.
 SQ SEQUENCE 98 AA; 10478 MW; 677A9271 CPG32;

Query Match 33.68, Score 82, DB 8, Length 98,
 Best Local Similarity 36.78; Pred. No. 1.97e+00;
 Matches 11; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

DB 15 SSLSGLHRTHLISALLCLESMSLMYAL 44
 QY 3 ASLNIIDQVDITGVLIVSSTIFCHYIAL 32

RESULT 12
 ID 052260 PRELIMINARY PRT, 245 AA.
 AC 052260;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, last annotation update)
 DE PLASMID PED208 SCREENING EXCLUSION SYSTEM
 DE (SEF) GENES TRAS AND TRAT. ENCODING TRATP (FRAGMENT).
 GN TRAD.
 OS unidentified bacterium.
 OC plasmid IncFV PRD208.
 OC Bacteria; environmental samples.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86233783.
 RA PINLAY B.H., PARANCHYCH W.:
 "Nucleotide sequence of the surface exclusion genes tras and trat from
 the IncF0 lac plasmid PRD208."
 RI J. Bacteriol. 166:713-721(1986).
 DR EMBL: M13465; AAA88376.1;
 KW Plasmid.
 FT NON-TER
 SQ SEQUENCE 245 AA; 28363 MW; C9A85EDC CRC32;

Query Match 33.68, Score 82, DB 2, Length 215,
 Best Local Similarity 31.08; Pred. No. 1.97e+00;
 Matches 9; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

DB 3 LKMFQINNIISYVIVGIVGIAVLIM 31
 QY 5 LNIIDQVDITGVLIVSSTIFCHYIAL 33

RESULT 13
 ID 064737 PRELIMINARY PRT, 386 AA.
 AC 064737;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, last sequence update)
 DT 01-NOV-1999 (TREMblrel. 12, last annotation update)
 DE HOOKLESS-1-LIKE PROTEIN.
 CN T27613.17.

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliopsida; eudicotyledons;
 OC Core eudicotyledons; Rosales; Eustosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN 111
 RP SEQUENCE FROM N.A.
 KC STRAIN-CV, COLUMBIA;
 RA FORNSLEY S.D., LIN X., KETCHUM K.A., GROSSY M.I., FRANKEN P.C.,
 SYDES S.M., FAUL S., MASON T.M., FRIDLAND A.P., ADAMS M.D.,
 RA SOMERVILLE G.P., VENTER J.C.:
 "Submitted (May-1998) to the EMBL/Genbank/Genbank databases."
 RI EMBL: AC004165; AAC16963.1;
 DR MEDLINE: 98460; Archib;1749;29669
 DR PFM: PF00583; Acetyltransferase 1.
 SQ SEQUENCE 386 AA; 43917 MW; 7848c776 CRC32;

Query Match 33.28, Score 81, DB 10, Length 386,
 Best Local Similarity 32.34; Pred. No. 2.08e+00;
 Matches 10; Conservative 13; Mismatches 6; Indels 2; Gaps 2;

DB 243 TKRLIEPAPLS-YLLITKVSLEFGNLSLLO 272
 QY 5 LNIIDQVDITGVLIVSSTIFCHYIAL 34

RESULT 14
 ID 030094 PRELIMINARY PRT, 469 AA.
 AC 030094;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, last sequence update)
 DT 01-AUG-1998 (TREMblrel. 07, last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN AFG143.
 OS Archaeoglobus fulgidus.
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 OC Archaeoglobus.
 RN 111
 RP SEQUENCE FROM N.A.
 RC STEIN VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE: 98049343.
 RA KLEIN H.-P., GLAYTON P.A., TOMR I.-P., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., JOHNSON P.J., GWINN M., HICKLEY P.K., PETERSON J.D.,
 RA RICHARDSON D.E., KERLAVAGE A.P., GRAHAM D.E., KYPILES N.C.,
 RA FLEISCHMANN P.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KIRPNESS E.F., TOSCHERTY R.A., MCKENNEY K., ATAMS M.D., LOFTUS B.,
 RA PETERSON S., REICH C.J., MCNEIL L.K., HANSEN I.H., SLIDER A., ZHOU L.,
 RA OVERBERGER P., GOGGAYNE J.D., WEILMAN J.P., MCGONALD J., UTTERBACK T.,
 RA GILSON M.H., SPRIGGS T., AKILACH P., KAIN B.P., STRES S.M.,
 RA SADOW P.W., FRANKEN P., FRANKEN P., FRANKEN P., FRANKEN P.,
 RA MASON T.M., OLSEN G.J., FRANKEN P.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.:
 "The complete genome sequence of the hyperthermophilic, sulphate-
 reducing archaeon Archaeoglobus fulgidus."
 RI Nature 396:364-370(1997).
 RI EMBL: AF001096; AAB91081.1;
 DR 1ick, AB0143.
 KW Hypothetical protein.
 SQ SEQUENCE 469 AA; 54027 MW; E085A8B CRC32;

Query Match 33.28, Score 81, DB 1, Length 469,
 Best Local Similarity 38.18; Pred. No. 2.08e+00;
 Matches 8; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

DB 31 FLTAIINLVEFLITFTGLFG 51
 QY 7 LNIIDQVDITGVLIVSSTIFCH 27

RESULT 15
 ID 045834 PRELIMINARY PRT, 503 AA.
 AC 045834;
 DT 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (ITEMBIO). 01, last sequence update)
 DT 01-NOV-1998 (ITEMBIO). 08, last annotation update)
 DE STARCH DEGRADING ENZYME PRECURSOR.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OT Clostridium.
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824.
 RA VERHAASSEL P.; VANDERLEVEN J.; VAN LEUVEN F.;
 RL Submitted (FEB 1993) to the EMBL/GenBank/CCDS databases.
 DR EMBL: X70334; CAA49800.1; -
 DR PIRAM: P00877; NIDC_P60.1.
 KW Signal.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 503 STARCH DEGRADING.
 SO SEQUENCE 503 AA; 53940 MW; 476DA117 CRC32;
 QUOTY Match 33.2%; Score 81; 108 2; Length 503;
 Host Local Similarity 38.5%; Prod. No. 2,680,00;
 Matches 10; Conservative 8; Mismatches 8; Indels 0; Gaps 0;
 ID 1 MKKFTNVLVGLVFIATVGHV 26
 ID 5 LNIHQQVDLGVLIIVSSITFGHY 30

Search completed: Fri Aug 25 16:02:47 2000
 Job time : 62 secs.

1
2
3
4


```

100 Mammalia: Eutheria: Primates: Catarrhini: Hylobatidae: Hylobates.
101 [1]
102 SEQUENCE FROM N.A.
103 MEDLINE: 98152416.
104 Vale A., Ortolan C., Richard F., Verrier P., Presse F., Schilling S.,
105 Dutilleul B., Nalon J.-L.
106 "Emergence of a brain-expressed variant melatonin receptor between
107 gene during higher primate evolution: a gene in search of a
108 function."
109 Mol. Biol. Evol. 15:196-214(1998).
110 -1- FUNCTION: THIS HORMONE INDUCES MELANOSOMAL AGGREGATION WITHIN
111 MELANOPHORES AND HAS AN ACTIVITY ANTAGONISTIC TO MELANOTROPIN.
112 -1- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATING HORMONE FAMILY.
113 -----
114 This SWISS-PROT entry is copyrighted. It is produced through a collaboration
115 between the Swiss Institute of Bioinformatics and the EMBL outstation
116 at the European Bioinformatics Institute. There are no restrictions on its
117 use by non-profit institutions as long as its content is in no way
118 modified and this statement is not removed. Usage by and for commercial
119 entities requires a license agreement (see http://www.isb-sib.ch/announc
120 or send an email to license@sib.ch).
121 -----
122
123 Query Match 42.4% Score 799 DB 1: Length 71
124 Best Local Similarity 47.5% Prod. No. 1.41e+00
125 Matches 97 Conservative 123 Mismatches 25 Gaps 1
126
127 DB 3 MNSSYLLITPFLSPSGI-LSA 25
128 12 VDLGYLVLSILFCHYIALAA 85
129
130 RESULT 5
131 ID MCHL_PANTR STANDARD PRT 71 AA.
132 AC 062689;
133 DT 15-JUL-1999 (rel. 38, created)
134 DT 15-JUL-1999 (rel. 38, last sequence update)
135 DT 15-FEB-2000 (rel. 39, last annotation update)
136 DE MELANIN-CONCENTRATING HORMONE PRECURSOR (PROMEL).
137 GN PMCH.
138 OS Pan troglodytes (chimpanzee).
139 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
140 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
141 [1]
142 KP SEQUENCE FROM N.A.
143 EX MEDLINE: 98152416.
144 Vale A., Ortolan C., Richard F., Verrier P., Presse F., Schilling S.,
145 Dutilleul B., Nalon J.-L.
146 "Emergence of a brain-expressed variant melatonin concentrating hormone
147 gene during higher primate evolution: a gene in search of a
148 function."
149 Mol. Biol. Evol. 15:196-214(1998).
150 -1- FUNCTION: THIS HORMONE INDUCES MELANOSOMAL AGGREGATION WITHIN
151 MELANOPHORES AND HAS AN ACTIVITY ANTAGONISTIC TO MELANOTROPIN.
152 -1- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATING HORMONE FAMILY.
153 -----
154 This SWISS-PROT entry is copyrighted. It is produced through a collaboration
155 between the Swiss Institute of Bioinformatics and the EMBL outstation
156 at the European Bioinformatics Institute. There are no restrictions on its
157 use by non-profit institutions as long as its content is in no way
158 modified and this statement is not removed. Usage by and for commercial
159 entities requires a license agreement (see http://www.isb-sib.ch/announc
160 or send an email to license@sib.ch).
161 -----
162
163 EMBL: AF029497; AF052501;

```

```

164 KW Signal, Cleavage on part of basic residues; Hormone.
165 KM Neuropeptide.
166 FT NON_TER 1
167 FT SIGNAL <1
168 FT CHAIN 21
169 FT NON_TER 71
170 SV SEQUENCE 71 AA: 7042 MW: 7934.4904700877 PKC4.
171
172 Query Match 42.4% Score 799 DB 1: Length 71
173 Best Local Similarity 47.5% Prod. No. 1.41e+00
174 Matches 97 Conservative 123 Mismatches 25
175
176 DB 3 MNSSYLLITPFLSPSGI-LSA 25
177 12 VDLGYLVLSILFCHYIALAA 85
178
179 RESULT 6
180 ID MCHL_PANTR STANDARD PRT 71 AA.
181 AC 062689;
182 DT 15-JUL-1999 (rel. 38, created)
183 DT 15-JUL-1999 (rel. 38, last sequence update)
184 DT 15-FEB-2000 (rel. 39, last annotation update)
185 DE MELANIN-CONCENTRATING HORMONE PRECURSOR (PROMEL).
186 GN PMCH.
187 OS Pan troglodytes (chimpanzee).
188 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
189 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
190 [1]
191 KP SEQUENCE FROM N.A.
192 EX MEDLINE: 98152416.
193 Vale A., Ortolan C., Richard F., Verrier P., Presse F., Schilling S.,
194 Dutilleul B., Nalon J.-L.
195 "Emergence of a brain-expressed variant melatonin concentrating hormone
196 gene during higher primate evolution: a gene in search of a
197 function."
198 Mol. Biol. Evol. 15:196-214(1998).
199 -1- FUNCTION: THIS HORMONE INDUCES MELANOSOMAL AGGREGATION WITHIN
200 MELANOPHORES AND HAS AN ACTIVITY ANTAGONISTIC TO MELANOTROPIN.
201 -1- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATING HORMONE FAMILY.
202 -----
203 This SWISS-PROT entry is copyrighted. It is produced through a collaboration
204 between the Swiss Institute of Bioinformatics and the EMBL outstation
205 at the European Bioinformatics Institute. There are no restrictions on its
206 use by non-profit institutions as long as its content is in no way
207 modified and this statement is not removed. Usage by and for commercial
208 entities requires a license agreement (see http://www.isb-sib.ch/announc
209 or send an email to license@sib.ch).
210 -----
211
212 EMBL: AF029496; AF052497;

```


4 SINITLOVDLTVLWISSLTHGYIA 31

RESULT 12
ID OPSB_PAT STANDARD PRT 346 AA.

AC Q64652;
01-NOV-1997 (Rel. 45, Created)
01-NOV-1997 (Rel. 45, Last sequence update)
15-JUL-1998 (Rel. 46, Last annotation update)
11 BLUE-SENSITIVE OPSIN (BLUE CONE PHOTORECEPTOR PIGMENT).
CN BCP.

OS Rattus norvegicus (Rat).
000 Eumastella; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
000 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
[1]

SEQUENCE FROM N.A.
FC STRAIN-SPRANNE DAWLEY; TISSUE-PITNEAL GLAND;
EX MEDLINE; 97227593.

FA Zhao X., Haeseleer F., Fariss R.N., Huang J., Haehr W., Mian A.H.,
Pajczewski K.;
"Molecular cloning and localization of rhodopsin kinase in the
mammalian pineal";

EL Vis. Neurosci. 14:225-232(1997).

000 -1- PINCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
MEDIATE VISION. THEY CONSIST OF AN APPOPROTEIN, OPSIN, COVALENTLY
LINKED TO CIS-RETINAL.

000 -1- SUBCELLULAR LOCATION: INTRACELLULAR MEMBRANE PROTEIN.
000 -1- TISSUE SPECIFICITY: THE THREE COLOR PIGMENTS ARE FOUND IN THE CONE
PHOTORECEPTOR CELLS.

000 -1- PIM: SOME OF ALL OF THE CARBOXYL-TERMINAL SER OF THE PESTIDES MAY
BE PHOSPHORYLATED.
000 -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
000 OPSIN SUBFAMILY.

000 This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See: <http://www.sdb.sil.ac.uk/00000000>
or send an email to license@sdb.sil.ch).

000 EMBL: U64972; AAB05931.1; -

000 GCRDB: GCR 1410; -

000 PRAM: PR00001; 7TmL1; 1.

000 PRINTS: PR00248; OPSIN.

000 PRINTS: PR00574; OPSINBLUE.

000 PROSITE: PS00247; G-PROTEIN_PPTPTP; 1.

000 PROSITE: PS00248; OPSIN; 1.

KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
phosphorylation; G-protein coupled receptor.

FT DOMAIN 1 31 EXTRACELLULAR
FT TRANSMEM 32 56 1 (POTENTIAL)
FT DOMAIN 57 68 CYTOPLASMIC
FT TRANSMEM 69 94 2 (POTENTIAL)
FT DOMAIN 95 108 EXTRACELLULAR
FT TRANSMEM 109 128 4 (POTENTIAL)
FT DOMAIN 129 147 CYTOPLASMIC
FT TRANSMEM 148 171 4 (POTENTIAL)
FT DOMAIN 172 197 EXTRACELLULAR
FT TRANSMEM 198 225 5 (POTENTIAL)
FT DOMAIN 226 247 CYTOPLASMIC
FT TRANSMEM 248 271 6 (POTENTIAL)
FT DOMAIN 272 279 EXTRACELLULAR
FT TRANSMEM 280 304 7 (POTENTIAL)
FT DOMAIN 305 346 CYTOPLASMIC
FT CARBOHYD 12 12 PROBABLE
FT DISULFID 105 182 POTENTIAL
FT BINDING 291 291 RETINAL CHROMOPHORE
SEQUENCE 346 AA; 30957 MW; 185CA227321917010 CIRC64;

Query Match 31.6%, Score 77, EB 2, Length 346

Post Local Similarity 46.4%; Pred. No. 2,70e+00;
Matches 13; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

Db 66 PLNIIIVNSIAGPLPCFES-VFTVFA 92
4 SINITLOVDLTVLWISSLTHGYIA 31

RESULT 13
ID FXSA_SFPA STANDARD PRT 139 AA.

AC P47148;
01-OCT-1994 (Rel. 30, Created)
01-OCT-1994 (Rel. 30, Last sequence update)
01-NOV-1997 (Rel. 35, Last annotation update)
000 FXSA PROTEIN (FRAGMENT).
CN FXSA.

OS Serratia marcescens.
000 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
000 Serratia.
[1]

SEQUENCE FROM N.A.
FC STRAIN-SR41;
EX MEDLINE; 95148727.

FA Omori K., Akatsuka H., Komatsubara S.;

000 "Construction of a versatile promoter analysis vector and its use for
analysis of the Serratia marcescens aspartate promoter region.";
000 plasmid 32:243-247(1994).

000 -1- SUBCELLULAR LOCATION: INTRACELLULAR MEMBRANE PROTEIN. INNER MEMBRANE
(POTENTIAL).

000 This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See: <http://www.sdb.sil.ch/00000000>
or send an email to license@sdb.sil.ch).

000 EMBL: D12252; AAA02517.1; -

000 GCRDB: GCR 139; 13996 MW; A897780614A54090 CIRC64;

000 PRAM: PR00001; 7TmL1; 1.

000 PRINTS: PR00248; OPSIN.

000 PRINTS: PR00574; OPSINBLUE.

000 PROSITE: PS00247; G-PROTEIN_PPTPTP; 1.

000 PROSITE: PS00248; OPSIN; 1.

KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
phosphorylation; G-protein coupled receptor.

FT DOMAIN 1 31 EXTRACELLULAR
FT TRANSMEM 32 56 1 (POTENTIAL)
FT DOMAIN 57 68 CYTOPLASMIC
FT TRANSMEM 69 94 2 (POTENTIAL)
FT DOMAIN 95 108 EXTRACELLULAR
FT TRANSMEM 109 128 4 (POTENTIAL)
FT DOMAIN 129 147 CYTOPLASMIC
FT TRANSMEM 148 171 4 (POTENTIAL)
FT DOMAIN 172 197 EXTRACELLULAR
FT TRANSMEM 198 225 5 (POTENTIAL)
FT DOMAIN 226 247 CYTOPLASMIC
FT TRANSMEM 248 271 6 (POTENTIAL)
FT DOMAIN 272 279 EXTRACELLULAR
FT TRANSMEM 280 304 7 (POTENTIAL)
FT DOMAIN 305 346 CYTOPLASMIC
FT CARBOHYD 12 12 PROBABLE
FT DISULFID 105 182 POTENTIAL
FT BINDING 291 291 RETINAL CHROMOPHORE
SEQUENCE 346 AA; 30957 MW; 185CA227321917010 CIRC64;

Query Match 31.6%, Score 76; DB 1; Length 139;
Post Local Similarity 40.0%; Pred. No. 3,71e+00;
Matches 8; Conservative 9; Mismatches 2; Indels 1; Gaps 1;



T.: Attiach, P.; Kating, R.P.; Sykes, S.M.; Sadov, P.M.;
D'Andrea, K.P.; Bowman, C.; Fujita, C.; Gariand, S.A.;
Mason, T.M.; Olsen, G.T.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Woese, C.R.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon *Archaeoglobus fulgidus*.
#cross-references MIMD:98049343
#accession G69267
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule-type DNA
##residues 1-469 ##label K16
##cross-references GR:AP01066; GR:AF00978; NID:52689419; PID:92650499;
TIGR:AF0143

SUMMARY #length 469 #molecular-weight 52627 #checksum 7785

Query Match 33.2% Score 81; DB 2; length 469;
Best Local Similarity 38.1% Pred. No. 2,936+00;
Matches 8; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

Db 31 FLITAINISVELLIFITGPG 51
27 7 TILGGVDRGVYLVISSTLPG 27

RESULT 10
ENTRY S31940 #type complete
TITLE starch-degrading enzyme
ORGANISM #formal_name Clostridium acetobutylicum
DATE 06-Jan-1998 #sequence_revision 06-Jan-1998 #text_change 09-Sep-1999
S31940
REFERENCE Vernasselli, P.; Vanderleyden, J.; Leuven, F.
#authors Submitted to the EMBL Data Library, February 1993
#accession S31940
##status preliminary
##molecule-type DNA
##residues 1-503 ##label VER
##cross-references EMBL:X70334; NID:944916; PID:944917
SUMMARY #length 503 #molecular-weight 53930 #checksum 9332

Query Match 33.2% Score 81; DB 2; length 503;
Best Local Similarity 38.5% Pred. No. 2,942+00;
Matches 10; Conservative 8; Indels 0; Gaps 0;

Db 1 MKKIFNVLVGCHVPSAIIYCHSV 26
27 5 LNIITGVVDVIGYLVISSTLPGHT 30

RESULT 11
ENTRY P2K9AC #type complete
TITLE photosystem II chlorophyll a-binding protein psbc precursor -
red alga (*Gyrodinium aureolum*) chloroplast
ORGANISM #formal_name Chloroplast *Gyrodinium aureolum*
DATE 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
S20954
S20954
REFERENCE Kesteven, U.; Mait, U.; Zetsche, K.
#authors Plant Mol. Biol. (1992) 18:777-780
#journal An equivalent to bacterial *ompR* genes is encoded on the
#title plastid genome of red algae.
#cross-references MIMD:92216053
#accession S20855
##status translation not shown
##molecule-type DNA
##residues 1-473 ##label KES
##cross-references EMBL:X6779; NID:911798; PID:9444601; PID:921281
REFERENCE S25084
#authors Mait, U.; Zetsche, K.

#journal Plant Mol. Biol. (1992) 19:1001-1010
#title A 16 kb small single copy region separates the plastid DNA
inverted repeat of the unicellular red alga *Gyrodinium aureolum*: physical mapping of the 1P-flanking regions and
nucleotide sequences of the psbc-psbc, rps16, 5S rRNA and
rpl21 genes.
#cross-references MIMD:92379254
#accession S25086
##molecule-type DNA
##residues 1-473 ##label MAI
##cross-references EMBL:X62578; NID:911798; PID:9444601; PID:921281
GENETICS
#gene psbc
#genome chloroplast
CLASSIFICATION #superfamily photosystem II chlorophyll a-binding protein
psbc
KEYWORDS #phylogeny; chloroplast; photosynthesis; photosystem II;
thylakoid
FEATURE #domain transit peptide (thylakoid) #status predicted
#label TRPY
15-473
#product photosystem II chlorophyll a-binding protein
#psbc #status predicted #label MAI
SUMMARY #length 473 #molecular-weight 52487 #checksum 4379

Query Match 32.4% Score 79; DB 1; length 473;
Best Local Similarity 37.0% Pred. No. 5,286+00;
Matches 10; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Db 157 MTWITGCHLILGCAFLVIVKAMPFG 183
QY 1 MDASINIIIGVVDVIGYLVISSTLPG 27

RESULT 12
ENTRY JC4096 #type complete
TITLE hypothetical 8.6K protein
ORGANISM #formal_name Cyanobacterium *synechococcus*
DATE 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
JC4096
JC4096
REFERENCE Sugita, M.; Luo, L.; Ohta, M.; Tadani, H.; Matsubayashi, T.;
#authors DNA Res. (1995) 2:71-76
#journal Genes encoding the group 1 intron-containing rRNA and
#title subunit L of NADH dehydrogenase from the Cyanobacterium
synechococcus PCC 6301
#accession JC4096
##status preliminary
##molecule-type DNA
##residues 1-74 ##label SUG
##cross-references DDBJ:D42186
SUMMARY #length 74 #molecular-weight 8561 #checksum 8601

Query Match 32.0% Score 78; DB 2; length 74;
Best Local Similarity 37.0% Pred. No. 7,056+00;
Matches 10; Conservative 8; Mismatches 8; Indels 1; Gaps 1;

Db 1 MIVVILIAALYIALAGAYLVVPAALY 27
QY 1 MDASINIIIGVVDVIGYLVISSTLPG 26

RESULT 13
ENTRY S65529 #type complete
TITLE NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) nrd protein -
Vibrio alginolyticus
ALTERNATE_NAMES NADH-ubiquinone oxidoreductase nrd protein
ORGANISM #formal_name *Vibrio alginolyticus*
DATE 29-Sep-1999
#cross-references MIMD:92379254
#accession S65529; S51018


```

REFERENCE      S05523
#authors      Hayashi, M.; Hirai, K.; Umemoto, T.
#journal      FEBS Lett. (1995) 363:75-77
#title        Sequencing and the alignment of structural genes in the nqr
                operon encoding the Na(+)-translocating NADH-quinone
                reductase from Vibrio alginolyticus.
#cross-references EMBL:U04364; NID:q683412; FID:d1008557; PID:q893414
#accession     S05523
#status        preliminary
#molecule_type DNA
#residues      1-210 #label HAY
#cross-references EMBL:U04364; NID:q683412; FID:d1008557; PID:q893414
#journal      S051013
#authors      Beattie, P.; Iano, K.; Bourne, R.W.; Iacchi, D.; Pich, P.P.;
                Ward, E.B.
#journal      FEBS Lett. (1994) 356:333-338
#title        Cloning and sequencing of four structural genes for the Na
                (+)-translocating NADH-ubiquinone oxidoreductase of Vibrio
                alginolyticus.
#cross-references EMBL:U04445
#accession     S051013
#status        preliminary
#molecule_type DNA
#residues      1-210 #label BFA
#cross-references EMBL:U04445; NID:q663268; PID:Q0A85479.1;
                PID:q63273

GENETICS
#gene          nqr A; nqrD
CLASSIFICATION #superfamily conserved hypothetical protein M1H88
KEYWORDS       NAD: oxidoreductase
SUMMARY        #length 210 #molecular weight 22602 #checksum 9319

Query Match      32.0%; Score 78; EB 2; Length 210;
Best Local Similarity 39.4%; Pred. No. 7,056+00;
Matches 13; Conservation 10; Mismatches 8; Indels 2; Gaps 2;

DB      80 ASIVIVWAV-TRKALYDYSKSL-SVAVN111 110
      111 112 113 114 115 116 117 118 119
QY      3 ASNIIIOGVDTGLVIVLSILFGHYIALAA 35

RESULT 14
ENTRY   C09212
#type complete
#description conserved hypothetical protein M1H81 - Methanobacterium
                thermotrophicum (strain Delta H)
#organism    Methanobacterium thermotrophicum
#date        05-jan-1997 #sequence-revision 05-dec-1997 #text-change
                05-jan-1998
ACCESSION C09212
REFERENCE AB09000
#authors      Smith, D.R.; Doucette-Stamm, L.A.; Delouppery, G.; Lee, H.;
                Dubois, J.; Aldredge, T.; Basurazadeh, R.; Blakey, D.;
                Cook, R.; Gilbert, R.; Harrison, D.; Hong, L.; Kealey, P.;
                Lamm, W.; Rothner, B.; Oiu, D.; Spadofora, R.; Velez, R.;
                Wang, Y.; Wiczkowski, J.; Gibson, R.; Jiwani, N.; Caruso,
                A.; Bush, D.; Sifer, H.; Patwell, D.; Prabhakar, S.;
                McDougall, S.; Shimer, G.; Coyall, A.; Petrokavski, S.;
                Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
                J.; Reeve, J.N.
#journal      J. Bacteriol. (1997) 179:7135-7155
#title        Complete genome sequence of Methanobacterium
                thermotrophicum delta H: functional analysis and
                comparative genomics.
#cross-references M01D:98047514
#accession     C09212
#status        preliminary; nucleic acid sequence not shown;
                translation not shown

#molecule_type DNA
#residues      1-343 #label MTH
#cross-references AB:AE008861; AB:AE006666; NID:q2921940; FID:q2621932
#experimental_source strain Delta H

GENETICS
#gene          M1H81

```

```

SUMMARY        #length 443 #molecular weight 42833
Query Match      42.0%; Score 78; DB 2; Length
Best Local Similarity 42.0%; Pred. No. 7,056+00;
Matches 8; Conservation 12; Mismatches 1;

DB      193 EMBL:U04364; S051013; S05523
QY      2 DASNIIIOGVDTGLVIVLSILFG 25

RESULT 15
ENTRY   S73303
#type complete
#description tryptophan synthase (EC 1.2.1.20) a-ph
                (porphyra purpurea) chloroplast
#organism    formal_homo chloroplast Porphyra purf
                19-Mar-1997 #sequence-revision 09-May
                16-Jul-1999
#date        16-Jul-1999
ACCESSION S73303
REFERENCE S73108
#authors      Keith, K.; Mathland, J.
#journal      Plant Mol. Biol. Rep. (1995) 13:433-43
#title        Complete nucleotide sequence of the fo
                chloroplast genome.
#accession     S73303
#status        isolated, not sequenced not shown
#molecule_type DNA
#residues      1-273 #label RE1
#cross-references EMBL:U04445; NID:q1276662; FID:
                PID:q127664;
                FID:q127664
#note         The nucleotide sequence was submitted
                to Genbank on 06-04-1995

GENETICS
#gene          trpA
CLASSIFICATION #superfamily tryptophan synthase alpha
KEYWORDS       synthase alpha chain homology
                carbon-oxygen rease; chloroplast; nqr
                biosynthesis
FEATURE        #domain tryptophan synthase alpha
                (trpA)
SUMMARY        #length 273 #molecular weight 29833
Query Match      41.4%; Score 77; EB 2; Length
Best Local Similarity 41.4%; Pred. No. 9,432+00;
Matches 11; Conservation 9; Mismatches 10;

DB      85 EMBL:U04364; S051013; S05523
QY      4 SNNIIIOGVDTGLVIVLSILFGHYIAL 44

Search completed: Fri Aug 25 16:00:44 2000
Job time : 28 secs.

```



00 OTHER INFORMATION: /note "Location of Intron 6"
01 FEATURE:
02 NAME/KEY: Peptide
03 LOCATION: 40..10
04 OTHER INFORMATION: /note "Tiretto leader sequence"
05 FEATURE:
06 NAME/KEY: Protein
07 LOCATION: 1..149
08 OTHER INFORMATION: /note "Factor Xa-Eight chain"
09 FEATURE:
10 NAME/KEY: Peptide
11 LOCATION: 143..194
12 OTHER INFORMATION: /note "Acidic Peptide"
13 FEATURE:
14 NAME/KEY: Protein
15 LOCATION: 195..448
16 OTHER INFORMATION: /note "Heavy Chain"
17 FEATURE:
18 NAME/KEY: Disulfide-bond
19 LOCATION: group(17..22, 50..61, 55..70, 72..81, 89..100, 96
20 LOCATION: ..109, 111..124, 132..192, 201..206, 221..237,
21 LOCATION: 460..364, 375..403)
22 SEQUENCE 488 AA; 5418 MW; 1232341 CN;
23
24 Query Match 40.7% Score 75; DB 1; Length 488;
25 Best Local Similarity 34.6% Pred. No. 3,67e+01;
26 Matches 9; Conservative 9; Mismatches 7; Indels 1; Gaps 1;
27
28 DB 1 MORPHOLINUSASLAG-ILLIGESLP 25
29 1..1111 1..1111 1..1111 1..1111
30 1 MDASLITLVOVDLTGILVISTLF 26
31
32 RESULT 5
33 ID US-08-691-071-4 STANDARD: PRT: 140 AA.
34 AN XXXXXX
35
36 Sequence 4, Application US/08691071
37 Patent No. 5789198
38
39 GENERAL INFORMATION:
40 APPLICANT: Akedion, Inc.
41 TITLE OF INVENTION: A NOVEL HUMAN LEPTIN
42 NUMBER OF SEQUENCES: 4
43 CORRESPONDENCE ADDRESS:
44 ADDRESSEE: Incyte Pharmaceuticals, Inc.
45 STREET: 3174 Porter Drive
46 CITY: Palo Alto
47 STATE: CA
48 COUNTRY: U.S.
49 ZIP: 94304
50 COMPUTER READABLE FORM:
51 MEDIUM TYPE: Diskette
52 OPERATING SYSTEM: DOS
53 SOFTWARE: FASTSEQ Version 1.5
54 CURRENT APPLICATION DATA:
55 APPLICATION NUMBER: US/08/691,071
56 FILING DATE: Filed Herewith
57 ATTORNEY/AGENT INFORMATION:
58 NAME: Billings, Lucy J.
59 REGISTRATION NUMBER: 36,749
60 TELEPHONE: 415-845-4166
61 TELECOMMUNICATION INFORMATION:
62 TELEPHONE: 415-855-0555
63 TELEFAX: 415-845-4166
64 INFORMATION FOR SEQ ID NO: 4:
65 SEQUENCE CHARACTERISTICS:
66 LENGTH: 140 amino acids
67 TYPE: amino acid
68 STRANDEDNESS: single
69 TOPOLOGY: linear
70 MOLECULE TYPE: peptide
71 IMMEDIATE SOURCE:
72 LIBRARY: GenBank
73 GI/NEI: 1197072
74 SEQUENCE 140 AA; 15663 MW; 109452 CN;
75
76 Query Match 29.9% Score 73; DB 1; Length 140;
77 Best Local Similarity 52.6% Pred. No. 5,14e+01;
78 Matches 19; Conservative 5; Mismatches 3; Indels 1; Gaps 1;
79
80 DB 22 GILVIVISCALPHNYPLP 40
81 1..1111 1..1111 1..1111 1..1111
82 16 CYLVIV SSTIPCHYIALH 33
83
84 RESULT 6
85 ID US-08-843-470-4 STANDARD: PRT: 140 AA.
86 AN XXXXXX
87
88 Sequence 4, Application US/08843370
89 Patent No. 5874535
90
91 GENERAL INFORMATION:
92 APPLICANT: Akedion, Inc.
93 TITLE OF INVENTION: A NOVEL HUMAN LEPTIN RECEPTOR
94 NUMBER OF SEQUENCES: 4
95 CORRESPONDENCE ADDRESS:
96 ADDRESSEE: Incyte Pharmaceuticals, Inc.
97 STREET: 3174 Porter Drive
98 CITY: Palo Alto
99 STATE: CA
100 COUNTRY: U.S.
101 ZIP: 94304
102 COMPUTER READABLE FORM:
103 MEDIUM TYPE: Diskette
104 OPERATING SYSTEM: DOS
105 SOFTWARE: FASTSEQ Version 1.5
106 CURRENT APPLICATION DATA:
107 APPLICATION NUMBER: US/08/843,370
108 FILING DATE: Filed Herewith
109 PRIOR APPLICATION DATA:
110 APPLICATION NUMBER: US/08/691,071
111 FILING DATE: August 1, 1996
112 ATTORNEY/AGENT INFORMATION:
113 NAME: Billings, Lucy J.
114 REGISTRATION NUMBER: 36,749
115 TELECOMMUNICATION INFORMATION:
116 TELEPHONE: 415-845-4166
117 TELEFAX: 415-845-4166
118 INFORMATION FOR SEQ ID NO: 4:
119 SEQUENCE CHARACTERISTICS:
120 LENGTH: 140 amino acids
121 TYPE: amino acid
122 STRANDEDNESS: single
123 TOPOLOGY: linear
124 MOLECULE TYPE: peptide
125 IMMEDIATE SOURCE:
126 LIBRARY: GenBank
127 GI/NEI: 1197072
128 SEQUENCE 140 AA; 15663 MW; 109452 CN;
129
130 Query Match 29.9% Score 73; DB 2; Length 140;


```

01 CC COUNTRY: USA
02 CC ZIP: 02173
03 CC COMPUTER READABLE FORM:
04 CC MEDIUM TYPE: Floppy disk
05 CC COMPUTER: IBM PC compatible
06 CC OPERATING SYSTEM: PC-DOS/MS-DOS
07 CC SOFTWARE: Patent Release #1.30
08 CC CURRENT APPLICATION DATA:
09 CC APPLICATION NUMBER: US/08/196,001
10 CC FILING DATE: 28-FEB-1995
11 CC CLASSIFICATION: 435
12 CC ATTORNEY/AGENT INFORMATION:
13 CC NAME: Graham, Patricia
14 CC REGISTRATION NUMBER: 32,227
15 CC REFERENCE NUMBER: MT-P49A2
16 CC TELECOMMUNICATION INFORMATION:
17 CC Telephone: 617-861-6240
18 CC Telefax: 617-861-9540
19 CC INFORMATION FOR SEQ ID NO: 4:
20 CC SEQUENCE CHARACTERISTICS:
21 CC LENGTH: 834 amino acids
22 CC TYPE: amino acid
23 CC TOPOLOGY: linear
24 CC MOLECULE TYPE: protein
25 CC SEQUENCE: 834 AA; 92828 MW; 1480242 CN;
26 SS
27 DB QUERY MATCH 29.5% SCORE 72; DI Z: Length 834;
28 DB Host Local Similarity 42.4%; Prod. No. 6,076,001;
29 UY Matches 14; Conservative 8; Mismatches 9; Indels 2;
30
31 RESULT 12
32 ID US-08-861-464-4 STANDARD: PRT: 834 AA.
33 XX xxxxxx
34 AC
35 DE Sequence 4, Application US/08861464
36 XX
37 CC Sequence 4, Application US/08861464
38 CC Patent No. 5874210
39 CC GENERAL INFORMATION:
40 CC APPLICANT: Quarente, Leonard P.
41 CC APPLICANT: Austriaco Jr., Nicotao
42 CC APPLICANT: Kennedy, Brian
43 CC TITLE OF INVENTION: Genes Determining Cellular Senescence
44 CC TITLE OF INVENTION: In Yeast
45 CC NUMBER OF SEQUENCES: 16
46 CC CORRESPONDENCE ADDRESSES:
47 CC ADDRESS: Hamilton, Brock, Smith & Reynolds, P.C.
48 CC STREET: Two Millenia Drive
49 CC CITY: Lexington
50 CC STATE: MA
51 CC COUNTRY: USA
52 CC ZIP: 02173
53 CC COMPUTER READABLE FORM:
54 CC MEDIUM TYPE: Floppy disk
55 CC COMPUTER: IBM PC compatible
56 CC OPERATING SYSTEM: PC-DOS/MS-DOS
57 CC SOFTWARE: Patent Release #1.0, Version #1.30
58 CC CURRENT APPLICATION DATA:
59 CC APPLICATION NUMBER: US/08/861,464
60 CC FILING DATE: 22-MAY-1997
61 CC CLASSIFICATION: 435
62 CC PRIOR APPLICATION DATA:
63 CC APPLICATION NUMBER: US 08/736,001
64 CC FILING DATE: 28-FEB-1995
65 CC PRIOR APPLICATION DATA:

```


01 APPLICATION NUMBER: P01/US94/09351
02
03 FILING DATE: 15-AUG-1994
04
05 PRIOR APPLICATION DATA:
06
07 APPLICATION NUMBER: US 08/107,408
08
09 FILING DATE: 16-AUG-1993
10
11 ATTORNEY/AGENT INFORMATION:
12
13 NAME: Grantham, Patricia
14
15 REGISTRATION NUMBER: 12,227
16
17 REFERENCE/DOCKET NUMBER: M17-6408A22
18
19 TELECOMMUNICATION INFORMATION:
20
21 TELEPHONE: 781-861-9540
22
23
24 INFORMATION CHAR SEQ ID NO: 4:
25
26 SEQUENCE CHARACTERISTICS:
27
28 LENGTH: 834 amino acids
29
30 TYPE: amino acid
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808

Query Match	29.5%	Score 72	DB 2	Length 834
Best Local Similarity	42.4%	Pred. NO.	6.07e+01	
Matches	14	Conserved	9	Mismatches 9

16 466 INASNNILTTTJLPTVNIWYIPRONQNYAL 197
:::|::| | : | : | : | : | :
27 1 MDASNITITQQVLE-TGYLLVISSIFCHYTAL 32

RESULT 13
US-08-477-451-8 STANDARD; PRI; 3200 AA.

XXXXXX

ID: Sequence B, Application US/08477451

Sequence 8, Application US/08477451

GENERAL INFORMATION:

05 APPLICANT: Coraack, Antonio
06 TITLE OF INVENTION: Holtrabacter pylori tag Region
07 NUMBER OF SPOUNCES: 46
08 CORRESPONDENCE ADDRESS:

ADDRESS: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville

COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

SOFTWARE: Patent In Review
 CURRENT: NOT FOR SALE

APPLICATION NUMBER: US/0
FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

NAME: McElung, Barbara J
REGISTRATION NUMBER: 34

REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION

TELEPHONE: 310-601-2208
TELEFAX: 310-655-3542

SEQUENCE CHARACTERISTICS:

TYPE: amino acid

TOPOLOGY: Linear

CC MOLECULE TYPE: PROTEIN
SQ SEQUENCE: 3200 AA; 278579.000; 12612128 CND;

Query Match	23.1%	Score 71	DB 2	Local	
Best Local Similarity	4.1%	100	100	100	
Matches	8	Conservative	4	Mismatches	4

10b	401	HHHJASVFECHWVET	417
	:	:	:
2y	17	YLVVISTFLGRVTAAL	84

RESULT	14		
ID	05 08 463 649 2	ADDRESS	1111 112 00

[illegible]

Sequence 2, April 1941 to May 1941

CC GENERAL INFORMATION:
CC APPLICANT: Shiqin, Liao, et al.
CC APPLICANT: Shiqin, Liao, et al.

CC APPLICANT: TAKASHIRO, Masaru, Ka
CC APPLICANT: Kaminuma, Miroku
CC APPLICANT: Yamada, Toru

CC APPLICANT: Kikuchi, Yoshio
CC APPLICANT: Emi, Shigeru
CC LETTER AND INFORMATION ON THE MATTER

CORRESPONDENCE ADDRESS:
ADDRESS: Fishburn, H.L.,
Address:

CITY: New York
STATE: New York

CC COMPUTER READABLE FORM
CC MEDIUM TYPE: C, ILLUMINATED KEY

```
CC SOFTWARE: Patent 11, Release 11 #
CC CURRENT APPLICATION DATA:
```

CC CLASSIFICATION: 43
ATTORNEY/AGENT INFORMATION:

TELECOMMUNICATIONS INTERNATIONAL
KIPPERING/DOCK: 11-0000-0000
CC
CC

TELEPHONE: (212) 419-4441
FAX: 236268
INFORMATION FOR SUPPLIERS

CC LENGTH: 332 amino acids
 CT TYPE: amino acid
 CC TOPOLOGY: linear

Query Match: 0.878 Score: 7

Best Local Similarity: 40.0%; Pred N
Matches: 13; covered: 1; not covered: 1

159 JOURNAL OF POST KEYNESIAN ECONOMICS

15 Jojoba
 16 M09515387-A.
 17 08-JUN-1995.
 18 FD 30-NOV-1994. m3686
 19 FR 30-NOV-1993. US-160602.
 20 P6 23-JUN-1994. US-285947.
 21 (CALI.) CALGENE INC.
 22 Landisbuhl KD, Cassner MW, Metz JG,
 23 WP1: 95-215267/28.
 24 D8 N-PDBE: 660208.
 25 PT production of very long chain fatty acid(s) in plant(s) - to produce
 26 drought and stress resistant transgenic plants)
 27 Example: Fig 3, 14pp; English.
 28 RNA is isolated from jojoba embryos and used to construct a cDNA
 29 library. Synthetic oligos are synthesized which overlap, so the
 30 AA sequences of selected peptide fragments and segments of DNA
 31 are produced. The DNA fragments and by PCR are labeled and used
 32 as a probe to screen clones from the cDNA libraries. An approx.
 33 1500 nt. jojoba cDNA clone is obtained. In this manner, Comparison
 34 to the peptide fragments in R7715-E77192 reveals the presence of
 35 each of these peptides in the translated sequence, with the
 36 exception of S01123. The mRNA is approx. 2kb. Preliminary DNA
 37 sequence of a jojoba gene is presented in G06107/R77164. Further
 38 DNA sequence analysis of additional clones indicates that there are
 39 at least 2 classes of cDNAs encoding this jojoba protein. The
 40 plastid contig. The entire coding region in pGEM7163 is constructed
 41 to contain a SalI site approx. 8 nt 5' to the ATG start codon and
 42 is designated pGCR7614. The complete DNA sequence of pGCR7614 is
 43 given in G06250/R77165. The motif differences between the two
 44 classes of cDNAs is the presence (G06707) or absence (G06209) of
 45 the 6 nt coding sequence for Asn 23 and 24 of R77164.
 46 Sequence 521 AA:

	Query Match:	37.7%	Score	92	DB Length	521	
	Best Local Similarity:	41.8%	Fold	Beta_4.23e+06			
Matches	13	Coverage	100	% Misclassification	5	Total Hits	1
Db	32 SINEHWK-CHYLSNALPVPFPG	59					
	1 11111111111111111111						
27	6 NIIQVCLCYLVELETETRMIALA	24					
RESULT	3						
ID	R79932	standard; Protein; 524 AA.					
AC	R79932;						
DE	16-JAN-1996 (first entry)						
SE	Jojoka wax synthase.						
KW	wax-synthase; long-chain alcohol-fatty-acid-transferase; wax ester; jojoka oilseed; transgenic plant; crop improvement; brassica. Simmondsia chinensis.						
OS	Location: ?nullifera						
FT	misc_difference 213						
FT	/note= "unidentified amino acid"						
PN	US545947-A.						
PI	29-AUG-1996						
PJ	20-NOV-1991; 796256.						
PR	20-MAY-1991; US 796256.						
PR	21-AUG-1992; US-933411.						
PK	13-NOV-1992; WO-930963.						
PR	20-MAY-1994; US-056299.						
PA	(CAL) CALGENE INC.						
PI	Lardizabal KD, Jassner MW, Metz JO;						
DR	WPI: 95-310894/40.						
OR	N-PDB: T04123.						
PT	DNA construct expressing jojoka wax synthase and fused to brassicin cells - useful for producing wax esters) for use in pharmaceuticals and cosmetics, etc						
PI	Claim 1: Column 47-50; 50pp; English						
PC	2 Wax synthases were isolated from a jojoka (<i>Simmondsia</i>) seedling progen, and tryptic fragments were sequenced (R79932). The sequence information was used to design oligonucleotide probes for screening jojoka embryo cDNA libraries. A 1.5 kb clone was isolated and additional sequences were identified by hybridization by reversegen. A						

CC 1733 bp clone was obtd. (T04123) that encoded wax synthase
SQ Sequence 524 AA;

Query Match	37.7%	Score 92	EB 1	Length 524
Best Local Similarity	44.8%	Prod No.	4,246,000	
Matches	13	Conservative	9	Mismatches 6; Indels 1; Caps 1;
DB	34	SINIMHVL-GYATLNAFLVPLIG	61	
6	NIHQVDLGLLVYSTLGHITALIA	34		

RESULT	4
11	K77164 standard; protein; 524 AA
12	177564

DE Very long chain fatty acid; fatty acyl-CoA

Pin	Key	Location/Qualifiers
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		
61		
62		
63		
64		
65		
66		
67		
68		
69		
70		
71		
72		
73		
74		
75		
76		
77		
78		
79		
80		
81		
82		
83		
84		
85		
86		
87		
88		
89		
90		
91		
92		
93		
94		
95		
96		
97		
98		
99		
100		

177
178

08-JUN-1995

PR 80-NOV-1993; US-150602

PA (CALT) CALGENE INC.

W11; 95-215-67/28.

Production of W

14. *Group and stress resistant transgene plant(s)*
15. *Example, Fig 2: 149pp; English.*

ably, synthetic oligos are synthesised which corresp. to the

are produced. The DNA fragments obtained by PCR are labeled and used

as a probe to screen clones from the cDNA libraries. All approx. 1500 nt. jobba cDNA clone is obt'd. in this manner. Comparison

each of these peptides in the translated sequence, with the

sequences of a tobacco cell is presented in Q90107/R77164. Further

DNA sequence analysis of additional clones indicates that there are at least 2 classes of cTNAs encoding this tobacco protein. The

plasmid only. The entire coding region in pGEM703 is constructed to contain a 5' start site already 8 nt' to the AUG start codon and

CC is designated pCGN7614. The complete DNA sequence of pCGN7614 is given in 090308/077165. The major difference between the two

classes of rRNA is the presence (19S) or absence (23S) of the 60S ribosomal subunit. The 19S and 23S rRNA sequences were determined by the method of Maxam and Gilbert (1980).

Sequence 524 AA;

Query Match 37.78, Score 92, DB 1, Length 524,

Matchless 13, Conservative 9, Misadventures 6, Indians 1, 60

Db 34 SINLIHVKL-GYHYLLISNALEFVFIPLLG 61

QY 6 NI11Q2VDTTGYILVISSILJCHYIALIA 34

ID K57486 standard, Protein, 524 AA.

DT 20-Sep-1993 (first entry)

Tryptic peptide; crop plant; fatty acid; fatty alcohol; acyltransferase

detergents; plastics; lubricants; pharmaceuticals; sperm whale.

OS	SYNCHRONIZED CIPHERS:
PH	KEY
	LOCATION/QUALIFIERS

was identified. 10N CFB. A cDNA library from RNA isolated from developing seeds of Lunaria was constructed, and 10N CFB was used to screen this library. Three classes of cDNA clones were isolated: Lunaria 1, Lunaria 5 and Lunaria 27. Lunaria 5 shares approx. 85% homology with the Brassica C220 clones, and it is suggested that it is active in seed oil fatty acid elongation.

Query Match 32.8% Score 80; DB 1; Length 498;
Best Local Similarity 29.0%; Pred. No. 3,227,017;
Matches 9; Conservative 14; Mismatches 7; Indels 1; Gaps 1;

DB 56 LDLSLQLPNVLVGFHFFIIFASIFVPM 96
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
QY 5 LNIILQ-OVDLGYLLVLSITLPGHYIALLA 34

RESULT 9
ID R77170; standard; Protein; 208 AA.

AC R77170;
DE 04-DEC-1995 (first entry)
DE Lunaria condensing enzyme clone 10N CFB.
KW Lunaria; condensing enzyme.

OS Lunaria annua.
PN K09515387-A.
PI 08-JUN-1995.
PI 30 NOV-1994; 013686.
PI 30 NOV-1994; US-160602.
PI 23-JUN-1994; US-265947.
PI (GALT) CALSENE INC.
PI Lardizabal KD, Lassarre MM, Maiz JB;
PI WP1: 95-255257/28.
PI N-PSDB; Q90216.

PI Production of very long chain fatty acid(s) in plant(s) is produced through drought and stress resistant transgenic plant(s) Example; Figure 11; 14pp; English.

PI The CE15 and CE20 Brassica cDNA sequences (see Q90210, Q90211 and Q90212) and the condensing enzyme encoding sequence from Jolyoba (Q90208) were used in determining primers Q90221-Q90225. From conserved AAs, these primers were variously used to PCR (RT-PCR) amplify fragments from RNA isolated from developing seeds of Lunaria annua, Tropaeolum majus (Nasturtium), and green illiquies of Arabidopsis thaliana. The primers most successfully utilized were Q90221 and Q90222. These primers were used to produce three clones encoding a portion of the condensing enzyme from Arabidopsis, from Lunaria a single clone was identified, 10N CFB. A cDNA library from RNA isolated from developing seeds of Lunaria was constructed, and 10N CFB was used to screen this library. Three classes of cDNA clones were isolated: Lunaria 1, Lunaria 5 and Lunaria 27. Lunaria 5 shares approx. 85% homology with the Brassica C220 clones, and it is suggested that it is active in seed oil fatty acid elongation.

Sequence 208 AA;

Query Match 31.1%; Score 76; DB 1; Length 208;
Best Local Similarity 29.0%; Pred. No. 6,366,017;
Matches 9; Conservative 13; Mismatches 8; Indels 1; Gaps 1;

DB 39 LDIISLQLPNVLVGFHFFIIFASIFVPM 69
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

QY 5 LNIILQ-OVDLGYLLVLSITLPGHYIALLA 34

RESULT 10
ID R42502; standard; Protein; 262 AA.

AC R42502;

DE 22-FEB-1994 (first entry)

DE A. rhizogenes NIAES1724 megaio-plasmid coded protein.

DE Dwarfing; rooting; plants; breeding; disease prevention;

DE Identification; classification; pathogens.

OS Arabidacterium rhizogenes NIAES1724.

PN J05192156-A.

PI 03-AUG-1993.

PI 10 DEC 1991, 360827.
PI 10 DEC 1991; JP-360827.
PI (TWO) SYOWA HAKKO KOSYO KA.
PI (SUMO) SUMITOMO METAL IND LTD.
PI WP1: 93-277471/35.

PI N-PSDB; Q46099.
PI DNA having specified base sequence relating to dwarfing and rooting of plants - is used for breeding and prevention of diseases of plants e.g. identification and classification of pathogens Example; Page 9-11; 19pp; Japanese.
PI The sequence is that of a protein encoded by DNA from an Arabidacterium rhizogenes NIAES1724 megaio plasmid relating to the dwarfing and rooting of plants.

Query Match 31.1%; Score 76; DB 1; Length 262;
Best Local Similarity 37.5%; Pred. No. 6,460,017;
Matches 9; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

DB 155 VEEDPESGFIAPSTLEPIHLEL 178
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
QY 9 LQVDLGYLLVLSITLPGHYIAL 32

RESULT 11
ID R42506; standard; Protein; 279 AA.

AC R42506;

DE 16-FEB-1994 (first entry)

DE Protein encoded by root-induction gene roih from A. rhizogenes.

DE hairy root induction; RI plasmid; T-DNA; plant cell.

OS Arabidacterium rhizogenes MAFR03-01724.

PN J05192154-A.

PI 03-AUG-1993.

PI 29 JAN 1992; 013058.

PI 29 JAN 1991; JP-0093385.

PI (DAIC) DAICEL CHEM IND LTD.

PI WP1: 93-277469/35.

PI N-PSDB; Q46113.

PI Hairy like root inducing gene present on T-DNA region of RI plasmid for artificially controlling induction of hairy type root in plant cell

PI Claim 3; Page 2; 17pp; Japanese.

PI The roih gene was localized to a 95bp NaeI SphI restriction fragment of A. rhizogenes 1 DNA. Three rol ("root induction") genes were isolated which can be used to control hairy root induction in plant cells. See also Q46112 and Q46114.

Sequence 279 AA;

Query Match 31.1%; Score 76; DB 1; Length 279;
Best Local Similarity 37.5%; Pred. No. 6,460,017;
Matches 9; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

DB 172 VEEDPESGFIAPSTLEPIHLEL 195
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
QY 9 LQVDLGYLLVLSITLPGHYIAL 32

RESULT 12
ID R14675; standard; Protein; 250 AA.

AC R14675;

DE 09-OCT-1991 (first entry)

DE Factor X-LACI hybrid protein.

DE Kunitz domain; blood; coagulation; inhibitor; Factor X;

DE Tissue factor; TF; Lipoprotein-Associated Coagulation Inhibitor.

DE Location/Qualifiers

FT protein

FT 1..171

FT /label= Xlc

FT peptide

FT 1..40

FT /label= prepro_leader

FT domain

FT 55..64

FT /label= CIA_domain

FT 89..150

FT domain

FT /label= growth_factor_domains

FI	domain	172..250
FI	disulfide_bond	/label: knitr_domain
ET	disulfide_bond	57..62
ET	disulfide_bond	90..101
ET	disulfide_bond	95..110
FI	disulfide_bond	112..121
FI	disulfide_bond	129..149
FI	disulfide_bond	146..149
ET	disulfide_bond	151..164
ET	disulfide_bond	189..206
ET	disulfide_bond	195..219
FI	disulfide_bond	211..242
FI	EP-49447-A	
FN	31-JUL-1991	
PF	21-JAN-1991	870008
PR	25-JAN-1990	US-470289
PA	(ONIV) ONLY OF WASHINGTON	
FI	Girard J, Broze GJ	
UR	MP1: 91-224839/31	
DR	N-PSDB: 012776	
PI	New factor X-LACI hybrid protein : comprises light chain of factor X and A1-1-155 heavy chain for use as anticoagulant	
PS	Disclosures: Page 12-14: 1pp; English	
CC	The protein is used as a blood coagulation inhibitor in mammals. It is believed to mimic the Xa/LACI complex in binding to and inhibiting VIIa/tissue factor. LACI inhibits via a novel feedback mechanism requiring degradation of Xa (a prod. of VIIa/TF activation); XI/LACI inhibits VIIa/TF activation directly.	
CC	The DNA allows prodn. of XI/LACI by introduction of the gene into cells suitable for expression, e.g. E. coli or CHO cells.	
CC	Sequence 250 AA:	

CC	Factor Xa to convert Fibrinogen to thrombin. An
CC	with Factor IX; are located near proline acids and
CC	See also RZ5512.
SO	Sequence 436 AA.

Query Match	Similarity	Score	E-Value
Host Local	97	Conservative	No. Matches
Matches	97	Conservative	No. Matches

Dn	1 MGRNVLISASLALGLLDELLE
QY	1 MASINILIQVNLIVLVLSGLF

RESULT	14
ID	R88468 standard; protein 3; AA
AC	R88468;
DI	30-AUG-1996 (first entry)
DE	Human two chain factor Xa1 single chain precursor
KW	human factor x; single chain factor x1; precurs
KW	factor Xa1; heparin peptide linker; Asn88; Ala185; pro
KW	not enzymatic activity; hemostasis regulator; f
KW	treatment; thrombus formation; vasculature patholo
KW	thrombin induced; restenosis; inflammation; proteo
KW	wound healing; bypass factors; replacement therapy
KW	imaging; diagnosis; antibody production.
OS	Homo sapiens.
IH	Key
FT	peptide
FT	/note "localization"
FT	/note "protein domain sequence"
FT	protein
FT	/note "heparin peptide linker"
F1	domain
F1	/note "enzymatic"
F1	/note "GMA domain"
FT	domain
FT	/note "high factor domains"
FT	misc_difference 101
F1	misc_difference 271
F1	/note "three as beta in specificity"
F1	misc_difference 308
FT	misc_difference 308
FT	/note "acid type Ser substr. acti
FT	disulfide_bond 56..99
F1	disulfide_bond 88..99
F1	disulfide_bond 97..125
F1	disulfide_bond 110..119
FT	disulfide_bond 127..138
FT	disulfide_bond 144..147
FT	disulfide_bond 149..162
FT	disulfide_bond 170..201
FT	disulfide_bond 190..195
FT	disulfide_bond 210..226
FT	disulfide_bond 339..353
FT	disulfide_bond 364..392
FN	W09600577-A1.
PN	11-JAN-1996.
PP	26-JUN-1995; 008468.
PR	25-JUN-1994; US-2080048.
FA	(CDDT-) GQR THERAPEUTIC INH
P1	Siba U, Woll DL:
DR	WPI: 96-077335/08.
P1	New pre-coagulant compositions for the treatment of
P1	modified blood factors selected from factor Aa, Fa
P1	Vlla and activated protein."
PS	Example 3; Fig 3; Gppp English.
CC	The present protein is the human factor X derived,
CC	factor X1, a precursor for the two chain factor X
CC	differs from factor X in that the factor X activat
CC	been replaced with a heparin-peptide linker, and the x

 RELEASE
 (TX)

Release 3.1A John F. Collins, BioComputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Aug 25 17:57:38 2000; Master time 9.40 Seconds

501,644 Million cell updates/sec

Tabular output not generated.

Title: >US-09-257-585-5

Description: (1-68) from US09257585.pcp

Perfect Score: 522

Sequence: 1 MALLINEPFLNSYLFVPI.....KKDKGKSGSLDFHGLPOL 68

Scoring table:

Gap 11

Scatched: 225878 seqs, 6933122 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

splemb112

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_humans

5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle

9:sp_phase 10:sp_plant 11:sp_protist 12:sp_unclassified

13:sp_vertebrate 14:sp_virus

Statistics: Mean 38.589; Variance 80.103; scale 0.482

Prod. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description	Prod. No.
1	94	18.0	361 7	MHC PDI MAJOR TRANSPIA	5,390-01
2	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
3	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
4	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
5	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
6	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
7	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
8	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
9	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
10	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
11	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
12	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
13	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
14	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
15	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
16	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
17	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
18	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
19	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01
20	94	18.0	364 7	MHC CLASS I ANTIGEN	5,390-01

Result ID	Score	Query Match Length	DB ID	Description	Prod. No.
21	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
22	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
23	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
24	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
25	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
26	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
27	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
28	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
29	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
30	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
31	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
32	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
33	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
34	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
35	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
36	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
37	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
38	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
39	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
40	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
41	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
42	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
43	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
44	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01
45	87	16.7	361 7	MHC CLASS I ANTIGEN	5,390-01

ALIGNMENTS

Result ID	Score	Query Match Length	DB ID	Description	Prod. No.
1	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
2	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
3	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
4	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
5	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
6	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
7	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
8	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
9	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
10	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
11	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
12	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
13	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
14	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
15	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
16	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
17	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
18	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
19	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01
20	94	18.0	361 7	MHC CLASS I ANTIGEN	5,390-01

PA GARNIER A., GREEN P., HAWKINS J., MILLER L., OJER M., JOHNSON L.,
 PA JONES M., KERSHAW J., KIRSHEN J., LAISIER N., LAFRIEL P.,
 PA LIGHTNING J., LLOYD G., MONTREAV A., MORTIMORE B., O'CALLAGHAN M.,
 PA PARRONS T., PERCY G., PITTEN I., WATKINS A., SANDERS D., SHAWNEEN R.,
 PA SWALLDOW N., SMITH A., SONNEMMER E., STADEN R., STOLSTON J.,
 PA THURRY MIBEL J., THOMAS K., VADINI M., VALCHAN K., WATKINSON R.,
 PA WATSON A., WEINSTOCK L., WILKINSON-SPIRAT J., WOLLMAN P.,
 K1 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*
 K2 *elegans*.
 K3 Nature 368:32-38(1994).
 K4 [2]
 K5 SEQUENCE FROM N.A.
 K6 MILLER N.
 K7 Submitted (29-V-1995) to the EMBL/GenBank/DDBJ databases
 K8 [3]
 K9 SEQUENCE FROM N.A.
 KA WATERSTON R.J.
 KB WATERSTON R.J.
 KC EMBL: U00414; AA081409; 1;
 KD SEQUENCE: 167 AA; 18813 MW; A1919A47 CRC32;

Query Match	17.4%	Score 91	DB 5	Length 167
Overall Similarity	51.9%	Prod. No.	1,32e+00	
Matches	14	Conservative	?	Mismatches 8
				Indels 2
				Gaps 2
1b	39	YVVCYVAVVLSIVGLGKWKNSK	63	
09	24	YVHVAFTTITLALFSLTQVVSOK	50	

RESULT	11	
ID	022568	PRELIMINARY
AC	022568:	PRO: 301 AA.
PT	01-NOV-1996	(TREMBLrel_01, created)
FT	01-NOV-1996	(TREMBLrel_01, last sequence update)
DE	01-JAN-1999	(TREMBLrel_09, last annotation update)
IE	1994.3	PROTEIN.
GN	1994.3	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;	
OR	Rhabditina; Rhabditoidea; Rhabditidae; Peloderini; Caenorhabditis.	
KN	11)	
RP	SEQUENCE FROM N.A.	
RA	WHITE S.;	
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBD databases.	
CH		

RP SEQUENCING FROM N.A.
RA MEDLINE: 94150718.
RA WILSON R. AINSWORTH P. ADRIEN H. B. BAYLES J. COLEMAN M.
RA HONEFIELD J. HURTON J. CONNELL M. COOPER I. COOPER J. COULSON A.
RA CRAYTON M. DEAR S. DE E. DEBRIK R. DAVELLO A. DILLI E.
RA FAFNER A. GREEN P. HAWKINS T. HILLIER E. HIER M. JARVISON E.
RA JONES M. KERSHAW J. KIRSTEN T. LAISTER N. LATRILLE P.
RA LIGHTNING J. LLOYD C. MCMURRAY A. MORIMORE H. O'CALLAGHAN M.
RA PARSONS J. PERRY C. PIERCE I. POOLPA A. SANDINES D. SHOMKHEIN R.
RA SMALDON N. SMITH A. SONNHAMMER E. STADEN R. SULLIVAN J.
RA THIERER MEG. J. THOMAS K. VAUDIN M. VANIGAN K. WATFORD P.
RA WATSON A. WEINSTEIN J. WIKINSON-SPEART T. WILLIAMS P.
RT 2.2 Mb of continuous nucleotide sequence from chromosome III of *C. elegans*."
RI Nature 358:42-48(1994).
RI EMBL: Z75549; CAA99415.1;
OR SEQUENCE: 001 AA: 14541 MW: 96204302 CRO22;

Query Match:	17.4%	Score 91	PR 5	Length 201
Host Local Similarity:	29.0%	Prod. No.	1328-02	
Matches	9	Conservative	15	Mismatches 5; Indels 2; Gaps 2
Db	172	IPPTFLPPEWVSQVSLNVAIVSFL	201	
	1	I I I I	
	5	I N F E T L S L E V P G I S T W H A V E F I L	L 4	

```

REMARK: 12
ID 019243 PRELIMINARY: PRI: 364 AA.
PR 01-JAN-1998 (TREMUREL: 05, Created)
DT 01-JAN-1998 (TREMUREL: 05, Last sequence update)
DT 01-NOV-1999 (TREMUREL: 12, Last annotation update)
DE MHC CLASS I ANTIGEN.
GN PC1A.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
NC Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
NC 11
KN
RP SEQUENCE FROM N.A.
FA Scott, VAN J. A., GOTTINGER H. F., SACHS D. H., EDGE A. S. B.,
KI J. Immunol. 159:0-0(1997)
DK EMBL: AE014004; MAB5335.1;
DK HSSP: P40685; IAYE.
FT TREMUREL: F552290; IG-MHC; 1.
DE PRAM: PR00047; Ig; 1.
DK PRAM: PR00129; MHC_I; 1.
KW MHC.
SD SEQUENCE 364 AA; 40791 MW; 18C101B8 CMC32;

```

Unity Match	17.48	Score 41	DB 7	Length 364
Post Local Similarity	34.28	Pred. No.	1,332/60	
Matches	122	Conservative	11	Ind. Is. 2
		Mismatches	11	Gaps 2
2b	465	VFVGGTIVTAVLVYVAVVAVVIVKRRSEKGG	340	
cy	18	VPLGSLVYVAVFTTIDALFTSLCKVAVSRKGG	55	

RESULT	13	PRELIMINARY:	PRT:	527 AA.
ID	025526			
AC	025526			
D1	01-JAN-1998 (trEMBLrel). 05, Created			
D1	01-JAN-1998 (trEMBLrel). 05, Last sequence update)			
D1	01-NOV-1998 (trEMBLrel). 08, Last annotation update)			
DE	ABSTRACT: O ACETYLATION PROTEIN (OAG).			
CH	HFE85.			
OS	Helicobacter pylori (Campylobacter pylori).			
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;			
OR	Helicobacter.			
RN	[1]			
RE	SEQUENCE FROM N.A.			

RX MEDLINE 92394467.
 KA TOMR J.-F., WHITE O., KRIEGLER A.R., CLAYTON K.A., SUTTON G.G.,
 KA LILJESANDER R.O., KETCHUM K.A., KERN H.-J., ELLIOTT S., HOPPEL P.B.A.,
 KA LINDNER K., QUAKENBUSH S., THOU L., KIRKINIS E.E., PETERSON S.,
 KA LEBTOS R., PETERSON P., ELLIOTT S., PHILLIP H.G., GELBER A.,
 KA KERNENY E., STEINBERG L.M., LEE N., ADAMS M.D., HIXEY E.R.,
 KA BEGG D.E., GAYNOR J.D., STEINBERG T.F., PETERSON J.D., KELLEY J.M.,
 KA COTTON M.D., WEILMAN J.M., POLL C., BOWMAN C., MATTHEY L., WALLIN E.,
 KA HAYES W.S., BERGDAVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 KA VINTER J.C.:
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*
 PT *pylori*." *Science* 288:539-547 (1997).
 DL EMBL: A0000596; AAD07902.1; -
 DR EMBL: A0000596; AAD07902.1; -
 EN IIR: H00855; -
 KW Hypothetical protein, 61-14 MW, BL27454 CDS27;
 C SEQUENCE 527 AA; 61-14 MW; BL27454 CDS27;

Query Match	17/48	Score 91	E- ⁻⁰²	Ident %	5.2%
Fast Local Similarity	34/38	Fold No.	1/42600		
Matches	12	Conservative	9	Mismatches	15
				Indels	1
				Gaps	1
Dh	429 ELKEYLYPLGNRYKELIVYENLL LVETLGGFW 362				
	: : : : : : : : :				
Qy	1 FLNMYLPVPIGSIYYVDIAVFTEEDMIFPSIGW 45				

RESULT 14

ID Q9YAV8 PRELIMINARY: FRT: 537 AA.

AC Q9YAV8; 01-NOV-1999 (JEMBL01, 12, Created)

D1 01-NOV-1999 (JEMBL01, 12, last sequence update)

D2 01-NOV-1999 (JEMBL01, 12, last annotation update)

DE 537AA LONG HYPOTHETICAL PROTEIN.

GN APE1895.

OS Acetopyrum pernix.

OC Archaea; Crenarchaeota; Acetopyrum.

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN-K1.

RX MEDLINE: 99310339.

KA KAWABAYASHI Y., HINO Y., HOJIKAWA H., YAMAZAKI S., NAKAYAMA Y.,

KA JIN-NO K., TAKAHASHI M., SERINE M., KABA S., ANEAI A., KOSUGI H.,

KA HOSONUMA A., FUKUI S., NAKAI Y., NISHIJIMA K., NAKAYAMA H.,

KA TAKAMURA M., MASUDA S., FUNAHASHI T., TANAKA T., KUDO H. Y.,

KA YAMAZAKI J., KISHIDA N., OGUCHI A., ABEI K., KUDOTA K., NAKAMURA Y.,

KA NOMURA N., SAKO Y., KIKUCHI H.;

KT *Complete genome sequence of an aerobic hyper-thermophilic

KT crenarchaeon, Acetopyrum pernix K1.;

ML DNA RES: 6:83-101(1999).

DR EMBL: AP000062; BAB0900.1.;

SQ SEQUENCE 537 AA: 54871 MW: 2340086 CRC32;

Query Match

Best Local Similarity: 40.0%; Prob. No. 1.32e+00;

Matches: 10; Conservative: 9; Mismatches: 5; Indels: 1; Gaps: 1;

DB 499 YFYALXLLASIFAVETIDFV 523

QY 9 FFLNSTLFLVLSIV-VQIAVFI 32

RESULT 15

ID 062006 PRELIMINARY: FRT: 1360 AA.

AC 062006;

D1 01-JUL-1997 (JEMBL01, 04, Created)

D2 01-JUL-1997 (JEMBL01, 04, last sequence update)

D3 01-NOV-1998 (JEMBL01, 08, last annotation update)

DE FRUIT FLY RETROTRANSPOSON RING DNA, COMPLETE SEQUENCE (FRAGMENT).

OS Drosophila simulans (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Phlebotomidae; Phlebotomidae; Phlebotomidae;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN-WHITE CHOCOLATE; TISSUE-WHITE BODY; TRANSPOSON NINJA;

RX MEDLINE: 96252483.

KA OGURA K., IAKUCHI S., NAKAYAMA I., YAMAMOTO M.;

KT *Molecular structure of the transposable element ninja in drosophila

KT simulans.;

KT Genes Genet. Syst. 71:1-8(1996).

DR EMBL: D81207; BAB19771.1.;

DR FLYBASE: FBgn0015169; DsimNinja.

DE PRAM: PF00098; zf-cchc.1.

FI NON-TER

SQ SEQUENCE 1360 AA: 150935 MW: 6750166 CRC32;

Query Match

Best Local Similarity: 28.6%; Prob. No. 1.32e+00;

Matches: 8; Conservative: 12; Mismatches: 8; Indels: 0; Gaps: 0;

DB 1275 GAVRAVHVIASAAFAAVAVVYI 1302

QY 21 GSIVVDIAVFILLIAIFSIQWKVS 48

Search completed: Fri Aug 25 17:58:48 2000
Job time: 70 secs.



(a) (b) (c) (d) (e)


```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

[illegible]

FT TRANSMEM 149 159 D (M4) (POTENTIAL)
 FT DOMAIN 160 168 CYTOPLASMIC (POTENTIAL)
 FT TRANSMEM 169 189 E (M5) (POTENTIAL)
 FT TRANSMEM 190 206 EXTRACELLULAR (POTENTIAL)
 FT TRANSMEM 209 229 F (M5A) (POTENTIAL)
 FT TRANSMEM 230 246 CYTOPLASMIC (POTENTIAL)
 FT TRANSMEM 247 257 G (M5B) (POTENTIAL)
 FT TRANSMEM 258 277 EXTRACELLULAR (POTENTIAL)
 FT TRANSMEM 278 296 H (M6) (POTENTIAL)
 FT TRANSMEM 299 307 CYTOPLASMIC (POTENTIAL)
 FT TRANSMEM 308 328 I (M7) (POTENTIAL)
 FT TRANSMEM 329 360 J (M8) (POTENTIAL)
 FT TRANSMEM 361 381 K (M9) (POTENTIAL)
 FT TRANSMEM 382 391 CYTOPLASMIC (POTENTIAL)
 FT TRANSMEM 392 412 L (M10) (POTENTIAL)
 FT TRANSMEM 413 429 EXTRACELLULAR (POTENTIAL)
 FT TRANSMEM 430 450 M (M11) (POTENTIAL)
 FT TRANSMEM 451 456 EXTRACELLULAR (POTENTIAL)
 FT TRANSMEM 459 479 N (M12) (POTENTIAL)
 FT TRANSMEM 480 809 CYTOPLASMIC (POTENTIAL)
 FT CARBOHYD 350 350 POTENTIAL
 SO SEQUENCE 809 AA: 90744 MW: 180004544 4087A6 780544
 Query Match 17.0% Score 89; LH 1; Length 809;
 Best Local Similarity 41.0%; Pred. No. 6,75e-01;
 Matches 18; Conservative 13; Mismatches 25; Indels 2; Gaps 2;

DB 453 VPKKRLKLT-AAIVVFVTVI-ICILIRLVHLLVWSSKKKGVANSTHRCPEED 508
 QY 10 IINSLVFLVHISIVYDAVFTLIDALFPGKWSKSKDKEGSLDTHQDLFQ 67
 RESULT 14
 ID NAH2.KAT1 STANDARD PRI: 813 AA.
 AC 148763;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE SODIUM/HYDROGEN EXCHANGER 2 (NAH2) (H4) (EXCHANGER 2) (NH2 2) (H7).
 GN SLC9A2 OR NHE2
 OS Rattus norvegicus (Rat)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;
 NC [1]
 RP SEQUENCE FROM N.A.
 KC TISSUE STOMACH;
 KX MEDLINE: 93280160;
 KA Wang Z., Glikowski J., Shull G.E.;
 ET Primary structure and functional expression of a novel
 ET gastric epithelial isoform of the rat Na/H exchanger."
 RL J Biol Chem 266(11):925-11928(1993)
 RN 121
 RP SEQUENCE OF 117-813 FROM N.A.
 KC STRAIN-SPRANG-DAMELEY TISSUE-SMALL INTESTINE;
 KX MEDLINE: 94248205;
 KA Collins T.F., Honda T., Krabel S., Solus N.M., Conary J.;
 KA Dubois R., Chisham P.K.;
 ET Molecular cloning, sequencing, tissue distribution, and functional
 ET expression of a Na/H exchanger (NHE-2)."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:3938-3942(1993).
 CC -1- FUNCTION: INVOLVED IN PH REGULATION TO ELIMINATE ACIDS GENERATED
 CC BY ACTIVE METABOLISM OR TO COUNTER ACTIVE ENVIRONMENTAL
 CC CONDITIONS. MAJOR PROTON EXTRUDING SYSTEM DRIVEN BY THE INWARD
 CC SODIUM ION CHEMICAL GRADIENT. PLAYS AN IMPORTANT ROLE IN SIGNAL
 CC TRANSDUCTION.
 CC -1- SUBCELLULAR LOCATION: INTERNAL MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS MAY BE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN SMALL INTESTINE, COLON, AND
 CC STOMACH, WITH MUCH LOWER LEVELS IN SKELETAL MUSCLE, KIDNEY, BLADDER,
 CC TESTIS, UTERUS, HEART, AND LUNG.
 CC -1- PTM: PHOSPHORYLATED (PHOSPHO).
 CC -1- SIMILARITY: BELONGS TO THE NAH(+) / H(+) EXCHANGER FAMILY.

CC -1- CATION: THE PROTON. TRANSPORT AND EFFLUX
 CC DOMAINS IN THE NAH(+) / H(+) EXCHANGER VARY AMON
 CC THIS SWISS-PROT entry is copyrighted. It is provided
 CC between the Swiss Institute of Bioinformatics and
 CC the European Bioinformatics Institute. There are
 CC use by non-profit institutions as long as they
 CC modified and this statement is not removed. These
 CC entities requires a license agreement (see EMBL/EBI
 CC or send an email to license@ebi.ac.uk).

DB 454 VPKKRLKLT-AAIVVFVTVI-ICILIRLVHLLVWSSKKKGVANSTHRCPEED 508
 QY 10 IINSLVFLVHISIVYDAVFTLIDALFPGKWSKSKDKEGSLDTHQDLFQ 67
 RESULT 15
 ID SYS_H0R040 STANDARD PRI: 425 AA.
 AC 051244;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE SERP-TYRA SYNTHETASE (SER 5.1.1.11) (SERINE 11 NA)
 GN SERP OR BB0226;
 OS Bacillus subtilis (B. subtilis) (B. subtilis)
 NC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
 NC [1]
 RP SEQUENCE FROM N.A.
 KC STRAIN-ATCC 35210 / 691;
 KX MEDLINE: 98065943;
 Query Match 17.0% Score 89; LH 1; Length 425;
 Best Local Similarity 41.0%; Pred. No. 6,75e-01;
 Matches 18; Conservative 13; Mismatches 25; Indels 2; Gaps 2;

 M32761 (M)

Release 3.1A, John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1995-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mseqch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: FRI Aug 25 17:56:25 2000; MasPar time 6.35 seconds
 Tabular output not generated. 505,405 Million cell updates/sec

Title: S0S-09-257-585-5
 Description: (1-68) from US09257585.ppt
 Sequence: 522
 1 MALLNNRFFLNSTYLVPL.....KKGDKQKQSLDHPHCLPOL 68

Scoring Table:

142086 seqs, 47172406 residues

Post-processing: Minimum Match 0.8
 Listing first 45 summaries

Database: p1r64
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 48.943; Variance 87.054; scale 0.447

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	99	19.0	873	2	S26690	glycoprotein b - porc	2,046-01
2	99	19.0	883	1	VBELL1	glycoprotein b precursor	2,046-01
3	94	18.0	864	2	I46604	MHC pD14 transplacit	8,400-01
4	94	18.0	866	2	I46603	MHC pD14a transplacit	8,400-01
5	94	18.0	883	1	VGERTS	glycoprotein b precursor	8,400-01
6	93	17.8	138	2	S33708	hypothetical protein	1,110-00
7	93	17.8	344	2	S14409	membrane protein, 37k	1,110-00
8	93	17.8	366	2	A60469	MHC class I histocomp	1,460-00
9	92	17.5	437	2	S24959	NADH dehydrogenase fu	1,460-00
10	92	17.5	451	2	C71403	hypothetical protein	1,920-00
11	91	17.4	167	2	T16454	hypothetical protein	1,920-00
12	91	17.4	412	2	S73218	preproprotein translocas	1,920-00
13	91	17.4	527	2	S64526	arginate O acetyltras	1,920-00
14	91	17.4	637	2	S72676	hypothetical protein	1,920-00
15	90	17.2	283	2	E20171	conserved hypothetical	2,530-00
16	89	17.0	466	2	JH0545	class I histocompatib	3,310-00
17	89	17.0	466	2	JH0547	class I histocompatib	3,310-00
18	89	17.0	466	2	JH0546	class I histocompatib	3,310-00
19	89	17.0	869	2	A46747	Na/H-exchanging pro	3,310-00
20	89	17.0	813	2	A46748	Na/H-exchanging pro	3,310-00
21	88	16.9	425	2	E20128	serine tRNA ligase	4,330-00
22	88	16.9	433	2	H71044	hypothetical protein	4,330-00
23	88	16.9	877	1	A25962	mitosis inhibitor, wee	4,330-00

24	87	16.7	186	2	A4444	T-cell, Src
25	87	16.7	805	2	BB644	class I b
26	87	16.7	503	2	S26690	glycoprotein b precursor
27	87	16.7	527	2	S64526	arginate O acetyltras
28	87	16.7	437	2	S24959	NADH dehydrogenase fu
29	86	16.5	472	2	A60469	MHC class I histocomp
30	85	16.3	191	2	A60469	MHC class I histocomp
31	85	16.3	191	2	A60469	MHC class I histocomp
32	85	16.3	191	2	A60469	MHC class I histocomp
33	85	16.3	191	2	A60469	MHC class I histocomp
34	85	16.3	191	2	A60469	MHC class I histocomp
35	84	16.1	471	2	A4444	T-cell, Src
36	84	16.1	460	2	S16308	hypothetical protein
37	84	16.1	439	2	S22845	NADH dehydrogenase fu
38	84	16.1	971	2	S67773	hypothetical protein
39	84	16.1	1872	2	S67773	hypothetical protein
40	83	15.9	433	2	S24959	NADH dehydrogenase fu
41	83	15.9	433	2	S24959	NADH dehydrogenase fu
42	83	15.9	1084	2	S1004	cellulose
43	83	15.9	1084	2	S1004	cellulose
44	83	15.9	1084	2	S1004	cellulose
45	83	15.9	1897	2	S67773	hypothetical protein

ACCESSIONS

RESULT 1
 ENTRY S26690 #type complete
 TITLE glycoprotein b precursor
 ORGANISM Homo sapiens
 DATE 14 Jun 1995
 #authors 1
 #journal S26690
 #REFERENCE S26690
 #cross-references M01192260, S26690

ACCESSIONS
 REFERENCE
 #authors 1
 #journal S26690
 #REFERENCE S26690
 #cross-references M01192260, S26690

CLASSIFICATION
 KEYWORDS
 SUMMARY

Query Match 19.0% E-value 99.10E-21
 Best Local Similarity 44.3% E-value 2.04E-01
 Matches 92 Conserved 162 Mismatches 112

Db 607 EMBL/GenBank/CCDS/TrEMBL/RefSeq
 4 LINEELINSLYLVPL...KKGDKQKQSLDHPHCLPOL 68

RESULT 2
 ENTRY VBELL1
 TITLE glycoprotein b precursor
 ORGANISM Homo sapiens
 DATE 16 Jun 1995

ACCESSIONS
 REFERENCE
 #authors 1
 #journal S26690
 #REFERENCE S26690

CLASSIFICATION
 KEYWORDS
 SUMMARY

##authors Dabrowski, J.; Piekarczyk, G.; Bartak, V.; Hattner, P.R.
##journal Science (1997) 276:771-774
##title Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes.
##cross-references M010:92458234
##accession S10708
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule-type DNA
##residues 1-138 ##label DAN
##cross-references EMBL:M87049; NID:q83656; PID:q148217
##note The nucleotide sequence was submitted to the EMBL Data Library, November 1992
REFERENCE S10742
##authors Omori, H.
##submission submitted to the EMBL Data Library, January 1993
##description physical map of the *ecfA* region of the *E. coli* chromosome.
##accession S10745
##status preliminary
##molecule-type DNA
##residues 1-138 ##label OHM
##cross-references EMBL:L02122; NID:q145576; PID:q145579
REFERENCE A64720
##authors Hartnett, E.R.; Plunkett, III, G.; Hsieh, C.A.; Perna, N.T.; Bartak, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Roda, C.K.; Mayhew, G.E.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goepfert, M.A.; Rose, D.D.; Mau, B.; Shao, Y.
##journal Science (1997) 277:1453-1462
##title The complete genome sequence of *Escherichia coli* K-12.
##cross-references M010:97426617
##accession C65196
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule-type DNA
##residues 1-138 ##label HAT
##cross-references CR:AF000457; CH:000096; NID:q29294; PID:q170257; M010:134816
##experimental-source strain K-12, substrain MG1655
GENETICS
##gene *yagG*
##map-position 86 min
SUMMARY #length 138 #molecular-weight 15764 #checksum 1118
Query Match 17.8% Score 93; DB 2; Length 133;
Best Local Similarity 28.1% Pident No. 111e+00;
Matches 9; Conservation 12; Mismatches 19; Labels 1; Gaps 1.
DB 76 LKDSPIYIKIVYVGLT-AVYIIISIIIMAEI 106
QY 4 LINEDFLINSYLPVIGSIYVDIVFTILD 35
RESULT 7
ENTRY S14409 #type complete
ENTRY S14409 membrane protein, 37k, precursor, chloroplast inner envelope
ENTRY S14409 spinach
ORGANISM #journal name *Spinacia oleracea* #common name spinach
DATE 21-Nov-1993 #sequence_revision 10 Nov 1995 #local_change 01-May-1998
ACCESSIONS S14409 S16986
REFERENCE S14409
##authors Ippes-Warinton, M.; Fischer, F.; Wocher, E.; Link, T.A.; Flieger, U.I.
##journal Eur. J. Biochem. (1991) 195:361-368
##title cDNA sequence and deduced amino acid sequence of the precursor of the 37-kDa inner envelope membrane polypeptide from spinach chloroplasts. Its transit peptide contains an amphiphilic alpha-helix as the only detectable structural element.
##cross-references M010:91146572
##accession S14409
##molecule-type mRNA

##residues 1-44 ##label LSL
##cross-references EMBL:M87049; NID:q29227; PID:q148217
REFERENCE S16986
##authors Black, M.A.; Tarr, J.; Rogers, B.
##journal FEBS Lett. (1993) 257:167-170
##title Purification and characterization of a convective protein.
##cross-references M010:914830
##accession S16986
##residues 119-137; X:119-141;236-240 ##label
GENETICS
##gene *hsc70*
CLASSIFICATION #superfamily to chaperone
KEYWORDS blocked amino acid endoplasmic
FEATURE 1-21
22-344 #protein membrane protein, 72k, 85%
118-219 #domain 1-21; #domain 219-344
SUMMARY #length 344 #molecular-weight 39000
Query Match 17.8% Score 93; DB 2; Length 133;
Best Local Similarity 28.1% Pident No. 111e+00;
Matches 9; Conservation 12; Mismatches 19
DB 308 LKDSPIYIKIVYVGLT-AVYIIISIIIMAEI 106
QY 9 FLENSYLPVIGSIYVDIVFTILD 49
RESULT 8
ENTRY A60369 #type complete
ENTRY A60369 MHC class I histocompatibility antigen precursor, HLA
ORGANISM #journal name *Human* #common name human
DATE 03-Mar-1993 #sequence_revision 17-Mar-1999
ACCESSIONS A60369
REFERENCE A60369
##authors Fritsch, W.; Fritsch, G.; Goldstein, H.
##journal New Biol. (1990) 2:1024-1033
##title Expression of a class I MHC transgene in tissue-specific regulatory
##cross-references M010:9149292
##accession A60369
##molecule-type DNA
##residues 1-44
##note The authors failed to translate at positions 22, 118, 301, 339 (GAG) as shown, have correct sequences
CLASSIFICATION #superfamily class I histocompatibility
KEYWORDS immunoglobulin domain
FEATURE 1-21
22-287 #domain 1-21; #domain 287-344
SUMMARY #length 344 #molecular-weight 39000
Query Match 17.8% Score 93; DB 2; Length 133;
Best Local Similarity 28.1% Pident No. 111e+00;
Matches 9; Conservation 12; Mismatches 19
DB 308 LKDSPIYIKIVYVGLT-AVYIIISIIIMAEI 106
QY 9 FLENSYLPVIGSIYVDIVFTILD 49
RESULT 9
ENTRY S14959 #type complete

```

TITLE      NADH dehydrogenase (deguanine) (E1.6.5.3) chain 4 -
ORGANISM   Crithidia oncopelti mitochondrion (8866)
DATE       40-Sep-1993 #sequence_revision 30 Sep 1993 #text_change
11-Jun-1994

ACCESSIONS S34959
REFERENCE   Muslow, P. A.; Bennett, A.; Owing, H. K.; Kolesnikov, A. A.
#authors   submitted to the EMBL data library, October 1990
#accession S34959
#molecule_type DNA
#residues 1437 #label MAS
#cross-references EMBL:X56915; NID:q12879; PIRN:CA59491.1; PID:q12881
GENETICS
#gene       ND4
#name       mitochondrion
#protein_code SC6
CLASSIFICATION
#superfamily NADH dehydrogenase (ubiquinone) chain 4
#membrane-associated complex; mitochondrion; NAD; oxidative
#phosphorylation; oxidoreductase; respiratory chain
KEYWORDS    #length 437 #molecular_weight 52018 #checksum 2805
SUMMARY

Query Match      17.6%  Score 92; 19.2; Length 437;
Best Local Similarity 34.4%  Pred. No. 146e+00;
Matches 11; Conservative 10; Mismatches 10; Indels 1; Gaps 1;

DB 365 GENE:SVLVEE-YVIVEAFVESSEVETL 365
QY 3 GENE:PIFMSYLVFVPGISVYDVAVFTL 34

ENTRY      C71403 #type complete
TITLE      Hypothetical protein Arabidopsis thaliana
ORGANISM   Arabidopsis thaliana #common_name mouse-ear
cress
#molecule_type cDNA
DATE       03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change
05-Dec-1998
ACCESSIONS C71403
REFERENCE   Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.;
Dean, C.; Borchamp, P.; Dicks, W.; Van Steveninck, M.;
Srikumar, W.; Drost, L.; Ridley, P.; Hughes, S. A.; Patel,
K.; Murphy, G.; Pittanelli, P.; Wedler, H.; Wedler, E.;
Wentworth, R.; Weitzmeyer, T.; Pohl, T. M.; Terry, N.;
Gleason, J.; Villarroel, R.; de Clerck, R.; Van Montagu, M.;
Lechamy, A.; Aubrey, S.; Gy, I.; Kreis, M.; Iao, N.;
Kewatani, T.; Hempel, S.; Koller, P.; Edlan, K. E.; Kiefer,
M.; Scheller, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.;
James, R.; Montfort, A.; Pons, A.; Pridmore, P.; Douka,
A.; Vouklatou, E.; Milioni, D.; Hatzopoulos, P.;
Piravittan, E.; Obermaier, H.; Hilbert, H.; Duesterhoft, A.;
Moore, T.; Jones, J. D. G.; Eneva, T.; Palme, K.; Benes, V.;
Reichman, S.; Ausorgo, W.; Cooke, R.; Berger, G.; Delcove,
M.; Voel, M.; Volkart, G.; Mewes, H. W.; Klotzel, S.;
Schubeler, G.; Chakwal, N.
#journal    Nature (1998) 391:485-488
#title      Analysis of 1.9 Mb of contiguous sequence from chromosome 4
of Arabidopsis thaliana.
#cross-references MIMD:98121113
#accession C71403
#status     preliminary; nucleic acid sequence not shown;
translational not shown
#molecule_type RNA
#residues 14361 #label REV
#cross-references GB:29745; NID:q2244747; PID:q327413; PID:q2244774
GENETICS
#map_position 40099,403845
SUMMARY    #map_position 1561 #molecular_weight 15113 #checksum 9792
Query Match      17.6%  Score 92; 19.2; Length 1661;
Best Local Similarity 32.4%  Pred. No. 146e+00;

```

```

Matches      11;   Conserved         13;   Mismatches       10;   Indels        0;   Gaps          0;
DB    1-127 LPSSAP1PISSTLDDISMSPACNSIPIAKSW 1360
      1-111-11111111111111111111111111111111
QY     12 INSTAYEPGIGSYVDVIAVFIIIDAIIPSGCW 45

RESULT      11
TITLE       #16454      #type complete
AUTHORS     hypotetical protein F53B3.5 - Caenorhabditis elegans
ORGANISM    #formal_name Caenorhabditis elegans
DATE        20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
20-Sep-1999

ACCESSIONS  TL6454
REFERENCE    Z18515

AUTHORS     Miller, N.
SUBMISSION  submitted to the EMBL Data Library, November 1995
DESCRIPTION  the sequence of C. elegans cosmid F53B3.
ACCESSION   TL6454
###status   preliminary; translated from cDNA/EMBL/DDDBJ
###molecule_type DNA
###residues EMBL:04014, NIN:q1065474; FID:q1065480;
           PIDN:AA81409.1; CESP:F53B3.5

GENETICS
#gene       CESP:F53B3.5
#features    45/11 76/13 186/3
#length 167 #molecular_weight 18813 #checksum 6308

Query Match
Best Local Similarity 51.9%; Fred.No. 1.92e+00;
Mismatch 14; Conservative 3; Mismatches 8; Indels 2; Gaps 2;

DB    19 YVAD1-VFVVIS-FIVGIQICWNKSAK 63
      11111111111111111111111111111111111111
QY     24 YVVIATVFITIIAIFPSISGWKWSRK 50

RESULT      12
ENTRY       S73218      #type complete
TITLE       proprotein translocase scv - rod alga (porphyra purpurea)
              chloroplast
DEFINITION  #literal_name Chloroplast Porphyra purpurea
            19-Mar-1997 #sequence_revision 0; Ka; 1997 #real_change
            20-Sep-1999
ACCESSIONS  S73218
REFERENCE    S73108
AUTHORS     Reith, M.; Munholland, J.;
            Giant Mol. Biol. Rep. (1995) 13:333-335
TITLE       Complete nucleotide sequence of the Porphyra purpurea
            chloroplast genome.
ACCESSION   S73218
##status    preliminary; nucleic acid sequence not shown;
            translation not shown

##molecule_type DNA
##residues  1-411 ##label RE1
##cross-references EMBL:U08804; NIN:q1276652; PINN:AAC08183.1;
                 P10;q1276763
##note       the nucleotide sequence was submitted to the EMBL Data
                 library, October 1995

GENETICS
#gene       scv
#feature    chloroplast
CLASSIFICATION
KEYWORDS    #superfamily proprotein translocase scv
            #chloroplast protein transport transmembrane protein
SUMMARY     #length 411 #molecular_weight 45660 #checksum 38

Query Match
Best Local Similarity 33.4%; Fred.No. 1.92e+00;
Mismatch 14; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

DB    1-269 LTOTLONTELLAVLDIEFGNSGLVMAVSLLFFSYPTSI 310
      11111111111111111111111111111111111111

```

07 1 MAGLINFDPISLVVHIAVFIILDAIPST 42

RESULT 14
ENTRY 664626 #type complete
TITLE aldimine O-acetyltransfer protein - Helicobacter pylori (strain 26695)
ORGNAM 09-Aug-1997 #sequence_revision 09 Aug 1997 #extl_change 10-Oct-1997

ACCESSIONS 664626
REFERENCE 664620
#authors Tomb, J.F.; White, O.; Klevavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.; Peterson, S.; Loftus, B.; Richardson, P.; Dodson, P.; Khajep, H.; Giesick, A.; McKerny, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, J.W.; Berg, D.E.; Gocayne, J.D.; Whiteback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Midden, J.M.; Fultz, G.; Bowman, C.; Wathey, L.; Walli, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.N.; Smith, H.O.; Fraser, C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen Helicobacter pylori.
#cross-references MIMD:97394467

ACCESSION 664626
#status preliminary: nucleic acid sequence not shown:
#molecule_type DNA
#residues 1527 #label 10M
#cross-references GB:AE000966; GB:AE000511; NID:95105244; PID:92418988; HIGR:HP0855

SUMMARY #length 527 #molecular_weight 61234 #checksum 6351

Query Match 17.4% Score 91; DB 2; Length 527;
Best Local Similarity 34.3% Prod. No. 1920-00;
Matches 12; Conservative 9; Mismatches 13; Indels 1; Gaps 1;
DB 329 FLKELYIPLEGNNKELLYNNLI LVFLIGFW 362
07 1 FLNSVIFVPIGSLVYVHIAVFIILDAIPSTCW 45

RESULT 14
ENTRY 672576 #type complete
TITLE hypothetical protein AEP1895 - Aeropyrum pernix (strain K1)
ORGNAM 20-Aug-1999 #sequence_revision 20-Aug-1999 #extl_change 20-Aug-1999
ACCESSIONS 672576
REFERENCE 672570
#authors Kawabuchi, Y.; Hiro, Y.; Horiwaga, H.; Yamazaki, S.; Hatakey, Y.; Jinno, K.; Takahashi, M.; Sekino, M.; Baba, S.; Aoki, A.; Kosugi, H.; Hasegawa, A.; Tokui, S.; Nagai, Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Koshida, N.; Oguchi, A.; Aoki, Y.; Kubota, K.; Nakamura, Y.; Nomura, N.; Sato, Y.; Kikuchi, H.
#journal DNA Res. (1999) 6:84-101
#title Complete genome sequence of an aerotolerant hyper-thermophilic Crepitarchaeon, Aeropyrum pernix K1.
#cross-references MIMD:99310339
#accession 672576
#molecule_type DNA
#status preliminary
#residues 1537 #label KAW
#cross-references ODB:AP000062; NID:95105244; PIDN:FAA80900.1;
#experimental_source strain K1

GENEIDCS
#gene
#genus
#species
#strain
#date
#time
#job

SUMMARY #length 527 #molecular_weight 61234

Query Match 17.4% Score 91; DB 2; Length 527;
Best Local Similarity 34.3% Prod. No. 1920-00;
Matches 12; Conservative 9; Mismatches 13; Indels 1; Gaps 1;
DB 329 FLKELYIPLEGNNKELLYNNLI LVFLIGFW 362
07 1 FLNSVIFVPIGSLVYVHIAVFIILDAIPSTCW 45

RESULT 15
ENTRY 670171 #type complete
TITLE conserved hypothetical integral membrane protein - Helicobacter pylori (strain 26695)
ORGNAM 09-Aug-1997 #sequence_revision 09 Aug 1997 #extl_change 10-Oct-1997

ACCESSIONS 670171
REFERENCE 670170
#authors Fraser, C.M.; Venter, J.C.; Clayton, R.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.; Peterson, S.; Loftus, B.; Richardson, P.; Dodson, P.; Khajep, H.; Giesick, A.; McKerny, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, J.W.; Berg, D.E.; Gocayne, J.D.; Whiteback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Midden, J.M.; Fultz, G.; Bowman, C.; Wathey, L.; Walli, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.N.; Smith, H.O.; Fraser, C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen Helicobacter pylori.
#cross-references MIMD:97394467

ACCESSION 670171
#status preliminary: nucleic acid sequence not shown:
#molecule_type DNA
#residues 1527 #label 10M
#cross-references GB:AE000966; GB:AE000511; NID:95105244; PID:92418988; HIGR:HP0855

SUMMARY #length 527 #molecular_weight 61234

Query Match 17.4% Score 91; DB 2; Length 527;
Best Local Similarity 34.3% Prod. No. 1920-00;
Matches 12; Conservative 9; Mismatches 13; Indels 1; Gaps 1;
DB 329 FLKELYIPLEGNNKELLYNNLI LVFLIGFW 362
07 1 FLNSVIFVPIGSLVYVHIAVFIILDAIPSTCW 45

RESULT 15
ENTRY 670171 #type complete
TITLE conserved hypothetical integral membrane protein - Helicobacter pylori (strain 26695)
ORGNAM 09-Aug-1997 #sequence_revision 09 Aug 1997 #extl_change 10-Oct-1997

ACCESSIONS 670171
REFERENCE 670170
#authors Fraser, C.M.; Venter, J.C.; Clayton, R.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.; Peterson, S.; Loftus, B.; Richardson, P.; Dodson, P.; Khajep, H.; Giesick, A.; McKerny, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, J.W.; Berg, D.E.; Gocayne, J.D.; Whiteback, T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Midden, J.M.; Fultz, G.; Bowman, C.; Wathey, L.; Walli, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.N.; Smith, H.O.; Fraser, C.M.; Venter, J.C.
#journal Nature (1997) 388:539-547
#title The complete genome sequence of the gastric pathogen Helicobacter pylori.
#cross-references MIMD:97394467

ACCESSION 670171
#status preliminary: nucleic acid sequence not shown:
#molecule_type DNA
#residues 1527 #label 10M
#cross-references GB:AE000966; GB:AE000511; NID:95105244; PID:92418988; HIGR:HP0855



 RELEASE 3.1A JON F. COLLINS, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

(1M)

Release 3.1A JON F. COLLINS, Biocomputing Research Unit,
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msrch_pj protein - protein database search using Smith-Waterman algorithm
 Run on: Fri Aug 25 17:59:05 2000; Msrch Time 4.02 seconds
 Tabular output not generated. 445,015 Million cell updates/sec

Title: US-09-257-585-5
 Description: (1.68) from US09257585.pep
 Perfect Score: 522
 Sequence: 1 MAGLIRIRITLNGRIIVPL.....RNDGNGGSHLWQSLHGL-58

Scoring table:
 PAM 150
 Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-issued
 1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:Backlist1

Statistics: Mean 25.744; Variance 107.081; scale 0.240

pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length DB	ID	Description	Pred. No.
1	88	16.9	4	PCT-US93-0	Sequence 266, Applicat
2	88	16.9	4	US-08-118-	Sequence 266, Applicat
3	88	16.9	4	US-08-118-	Sequence 266, Applicat
4	87	16.7	2	US-08-284-	Sequence 25, Applicat
5	87	16.7	2	PCT-US93-0	Sequence 25, Applicat
6	87	16.7	2	US-08-417-	Sequence 25, Applicat
7	87	16.7	2	PCT-US92-0	Sequence 25, Applicat
8	84	16.1	4	US-08-118-	Sequence 268, Applicat
9	84	16.1	4	PCT-US93-0	Sequence 268, Applicat
10	84	16.1	4	US-08-118-	Sequence 268, Applicat
11	84	16.1	4	US-08-118-	Sequence 268, Applicat
12	84	16.1	4	US-08-118-	Sequence 268, Applicat
13	84	16.1	4	US-08-118-	Sequence 268, Applicat
14	84	16.1	4	US-08-118-	Sequence 268, Applicat
15	84	16.1	4	US-08-118-	Sequence 268, Applicat
16	84	16.1	4	US-08-118-	Sequence 268, Applicat
17	84	16.1	4	US-08-118-	Sequence 268, Applicat
18	84	16.1	4	US-08-118-	Sequence 268, Applicat
19	84	16.1	4	US-08-118-	Sequence 268, Applicat
20	84	16.1	4	US-08-118-	Sequence 268, Applicat
21	84	16.1	4	US-08-118-	Sequence 268, Applicat
22	84	16.1	4	US-08-118-	Sequence 268, Applicat
23	84	16.1	4	US-08-118-	Sequence 268, Applicat

24	75	14.4	4	US-08-118-	Sequence 266, Applicat
25	75	14.4	4	US-08-118-	Sequence 266, Applicat
26	74	14.2	4	US-08-118-	Sequence 266, Applicat
27	74	14.2	4	US-08-118-	Sequence 266, Applicat
28	74	14.2	4	US-08-118-	Sequence 266, Applicat
29	74	14.2	4	US-08-118-	Sequence 266, Applicat
30	74	14.2	4	US-08-118-	Sequence 266, Applicat
31	74	14.2	4	US-08-118-	Sequence 266, Applicat
32	74	14.2	4	US-08-118-	Sequence 266, Applicat
33	74	14.2	4	US-08-118-	Sequence 266, Applicat
34	74	14.2	4	US-08-118-	Sequence 266, Applicat
35	74	14.2	4	US-08-118-	Sequence 266, Applicat
36	74	14.2	4	US-08-118-	Sequence 266, Applicat
37	74	14.2	4	US-08-118-	Sequence 266, Applicat
38	74	14.2	4	US-08-118-	Sequence 266, Applicat
39	74	14.2	4	US-08-118-	Sequence 266, Applicat
40	72	13.8	4	US-08-118-	Sequence 266, Applicat
41	72	13.8	4	US-08-118-	Sequence 266, Applicat
42	72	13.8	4	US-08-118-	Sequence 266, Applicat
43	72	13.8	4	US-08-118-	Sequence 266, Applicat
44	72	13.8	4	US-08-118-	Sequence 266, Applicat
45	72	13.8	4	US-08-118-	Sequence 266, Applicat

A. COMMENTS

RESULT 1
 ID PCT-US93-08528 266 SIAM APLA PRT 4 AA
 AC: AAAAAA
 DT
 XX Sequence 266, Application 67/1059408528
 DE
 XX GENERAL INFORMATION:
 CC APPLICANT: New York University
 CC TITLE OF INVENTION: PEPTIDES OF G-05-0000
 CC FIELD OF INVENTION: PEPTIDES, AND COMBINATION
 CC NUMBER OF SEQUENCES: 40
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESS: 419 Second Street, N.W., Suite 100
 CC CITY: Washington
 CC STATE: D.C.
 CC COUNTRY: USA
 CC ZIP: 20004
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: 11.44 disk
 CC COMMENTS: 100% of sequence
 CC ORIGINATING SYSTEM: 100% of sequence
 CC SOFTWARE: Patent to Bioscience, 1.0, Version 1.0
 CC CURRENT APPLICANT: N.Y.U.
 CC ATTORNEY: N.Y.U.
 CC FILING DATE: 10/28/1992
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 08/07944236
 CC FILING DATE: 10/28/1992
 CC NAME: Townsend, Keith
 CC REGISTRATION NUMBER: 4000
 CC PCT PRIORITY: 08/07944236
 CC PCT PRIORITY: 08/07944236
 CC TELEPHONE: 202 628 1327
 CC TELEFAX: 202 628 1327
 CC INFORMATION FOR SEQ ID NO: 266:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 42 amino acids
 CC TYPE: amino acid
 CC STANDARDS: Single
 CC TOPOLOGY: linear

[illegible]

TELEX: 200154
 INFORMATION FOR SEQ ID NO: 25:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 182 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: amino acid
 SEQUENCE: 182 AA; 20469 MW; 157982 CN;

Query Match 16.7% Score 87; DB 2; Length 182;
 Best Local Similarity 45.7% Pred. No. 7,00e+00;
 Matches 10; Conservative 12; Mismatches 5; Indels 1; Gaps 1;

ID 115 ISGHFAFVLSFVLAAGVFIAADGV 142
 12 LNSYLFVPIGSIYVDIAV-FTLIDAI 38

RESULT 7
 ID PCT-US92 01785 25 STANDARD; PRT 182 AA.

AC XXXXX

Sequence 25, Application PC/TUS9201785

Sequence 25, Application PC/TUS9201785

GENERAL INFORMATION:

APPLICANT: The General Hospital Corporation

TITLE OF INVENTION: Re-direction of Cellular Immunity by Receptor

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB

COMPUTER: IBM PC/2 Model 502 or 553X

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: Wordperfect (Version 5.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/01785

FILING DATE: 1992/03/06

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/665,961

FILING DATE: March 7, 1991

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 40,162

REFERENCE/DOCKET NUMBER: 00766/119092

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 182 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: amino acid

SEQUENCE: 182 AA; 20469 MW; 157982 CN;

Query Match 16.7% Score 87; DB 4; Length 182;
 Best Local Similarity 45.7% Pred. No. 7,00e+00;
 Matches 10; Conservative 12; Mismatches 5; Indels 1; Gaps 1;

ID 115 ISGHFAFVLSFVLAAGVFIAADGV 142

12 LNSYLFVPIGSIYVDIAV-FTLIDAI 38

RESULT 8
 ID US-08-118-270-268 STANDARD; PRT 46 AA.
 AC XXXXX

Sequence 268, Application US/08118270

Sequence 268, Application US/08118270

Patent No. 5508384

GENERAL INFORMATION:

APPLICANT: Murphy, Randall B.

TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN

RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF

NUMBER OF SEQUENCES: 348

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWN AND NEWMARK

STREET: 419 Seventh Street, N.W., Suite 400

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/118,270

FILING DATE: 09-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/943,246

FILING DATE: 10-SEP-1992

ATTORNEY/AGENT INFORMATION:

NAME: Townsend, Kevin G.

REGISTRATION NUMBER: 34,033

REFERENCE/DOCKET NUMBER: MURPHY-2A

TELEPHONE: 202-628-5197

TELEFAX: 202-737-4528

TELEX: 446033

INFORMATION FOR SEQ ID NO: 268:

SEQUENCE CHARACTERISTICS:

LENGTH: 46 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Peptide

SEQUENCE: 46 AA; 5091 MW; 33173 CN;

Query Match 16.1% Score 84; DB 1; Length 46;
 Best Local Similarity 46.2% Pred. No. 1,21e+01;
 Matches 12; Conservative 9; Mismatches 3; Indels 2; Gaps 2;

ID 19 IFNMAPLPGSIAGSEFYI-PIAVIII 43
 10 IF-LNSYLFVPIGSIYVDIAVFTL 34

RESULT 9
 ID PCT-US94 08528 268 STANDARD; PRT 46 AA.
 AC XXXXX

Sequence 268, Application PC/TUS9408528

XX Sequence 248, Application PG/US9408528
 CC GENERAL INFORMATION:
 CC APPLICANT: New York University
 CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
 CC TITLE OF INVENTION: RECEPTORS AND COMPOSITIONS AND METHODS THEREOF
 CC NUMBER OF SEQUENCES: 348
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: BROWDY AND NEWMARK
 CC STREET: 419 Seventh Street, N.W., Suite 300
 CC CITY: Washington
 CC STATE: D.C.
 CC COUNTRY: USA
 CC ZIP: 20004
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: PCT/US94/08528
 CC FILING DATE: 09-SEP-1994
 CC PRIORITY APPLICATION DATA:
 CC APPLICATION NUMBER: US 07/943,246
 CC FILING DATE: 10-SEP-1992
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Townsend, Kevin G.
 CC REGISTRATION NUMBER: 34,033
 CC REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 202-628-5197
 CC TELEFAX: 202-747-4528
 CC TELEX: 248633
 CC INFORMATION FOR SEQ. 19 NO: 268:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 46 amino acids
 CC TYPE: amino acid
 CC STANDARDS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC SEQUENCE 46 AA: 5091 MW: 13151 CN:
 CC
 CC Query Match 16.1% Score 84; DB 4; Length 46;
 CC Host Local Similarity 46.2%; Pred. No. 1216-01;
 CC Matches 12: Conservative 9; Mismatches 4; Indels 2; Gaps 2;
 DB 19 IPNMAFLFSLGSEYI-PLAVLIL 43
 YY 10 IF-INSYLFVPGSYVVDIAVFIIL 34
 CC
 CC RESULT 10
 CC ID US-08-928-692-20 STANDARDS: PRT, 371 AA.
 CC
 CC XXXXXX
 CC Sequence 20, Application US/94042892
 CC Patent No. 598727
 CC GENERAL INFORMATION:
 CC APPLICANT: Freedy, Howard
 CC APPLICANT: Yaver, Deborah S.
 CC APPLICANT: Lassa, Michael
 CC APPLICANT: Hanson, Kim
 CC TITLE OF INVENTION: Methods for Modifying the Production of
 CC TITLE OF INVENTION: a Polypeptide
 CC NUMBER OF SEQUENCES: 60
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: NO. 598727, No. 598727, 1st of No. 598727th America, Inc.
 CC STREET: 405 Lexington Avenue

CC CITY: New York
 CC STATE: NY
 CC COUNTRY: USA
 CC ZIP: 10174
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Diskette
 CC COMPUTER: IBM Compat. 486
 CC OPERATING SYSTEM: DOS
 CC SOFTWARE: FastSeq for Windows Version 2.0
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/94/042892
 CC FILING DATE: 12 SEP 1994
 CC CLASSIFICATION: 483
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Lambitis, Elias J.
 CC REGISTRATION NUMBER: 34,718
 CC REFERENCE/DOCKET NUMBER: 4944-200 US
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 212-867-0124
 CC TELEFAX: 212-878-9476
 CC INFORMATION FOR SEQ. 11 NO: 2
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 371 amino acids
 CC TYPE: amino acid
 CC STANDARDS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: No. 598727
 CC SEQUENCE 371 AA: 41551 MW: 770016 CN:
 CC
 CC Query Match 16.1% Score 84; DB 4; Length 46;
 CC Host Local Similarity 46.2%; Pred. No. 1216-01;
 CC Matches 14: Conservative 14; Mismatches 10;
 DB 249 ILPLASTLEINLNVNGLAR VYSNL EIE-EIN-WMYERS
 YY 9 FPIHSTLVPIHSTLV VYLVAVLITLAVHSTSLWVYSHH
 CC
 CC RESULT 11
 CC ID US-08-619-554-2 STANDARDS: PRT, 1870 AA
 CC
 CC XXXXXX
 CC Sequence 2, Application US/94042892
 CC Patent No. 5821354
 CC GENERAL INFORMATION:
 CC APPLICANT: JOURNALS, Cambridge M.
 CC APPLICANT: CHERRET, Gary D.
 CC APPLICANT: CLEMAS, Joseph
 CC APPLICANT: EL-SHERPIEN, Mohammed
 CC APPLICANT: FOKK, Fredrick
 CC APPLICANT: KAHN, Donald
 CC APPLICANT: KELLY, Rosamund
 CC APPLICANT: MARIYAN, John
 CC APPLICANT: PAREN, S.A.
 CC APPLICANT: WATIN, Gary
 CC APPLICANT: WILSH, David
 CC APPLICANT: WILSH, David
 CC TITLE OF INVENTION: SYNTHESIS OF
 CC TITLE OF INVENTION: SYNTHESIS OF
 CC NUMBER OF SEQUENCES: 2
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: JOSEPH A. CHERRET, MEYER & CO.
 CC STREET: 126 EAST 10TH STREET, FLOOR 10, N.Y.C.
 CC CITY: NEW YORK
 CC STATE: NY
 CC COUNTRY: USA
 CC ZIP: 10003
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Diskette
 CC MODIFIER: IBM Compat. 486

CC MOLECULE TYPE: PROTEIN
 CC DESCRIPTION: PROTEIN
 CC ORIGINAL SOURCE:
 CC ORGANISM: Human
 CC FEATURE: Protein Kinase; Table 8 Column 43
 CC PUBLICATION INFORMATION:
 CC AUTHOR:
 CC AUTHORS: Banks, S. K.
 CC AUTHORS: Quinn, A. M.
 CC AUTHORS: Hunter, T.
 CC TITLE: The protein kinase family
 CC JOURNAL: Science
 CC VOLUME: 241
 CC PAGES: 42-52
 CC DATE: 1988
 CC SEQUENCE: 269 AA; 40511 MW; 469255 CH;
 Query Match: 15.5% Score 91; 108 Z; Length 269;
 Host Local Similarity: 28.6%; Prod. No. 2,090,01;
 Matches: 12; Conservative: 9; Mismatches: 20; Indels: 1; Gaps: 1;
 DB 175 LANLDRPADWSLSTVEANANIVLPDNGSWKLEPSD 216
 YY 12 INSLVPLGSLVIVHAFVHDAIPPSIGQWKSKECD 52
 RESULT 14
 ID US-08 485 590A-18 STANDARD: FBI: 69 AA.
 AC XXXXXX
 DE Sequence 18, Application US/08485590A
 XX Patent No. 5824861
 CC GENERAL INFORMATION:
 CC APPLICANT: Aldwinckle, Herbert S.
 CC APPLICANT: No. 5824861-111, John L.
 CC TITLE OF INVENTION: TRANSFERRING POLYMERS FROM WITH FIVE
 CC NUMBER OF SEQUENCES: 28
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESS: NIXON, HERBERT, DOWNEY & DOYLE LLP
 CC STREET: Clinton Square
 CC CITY: Rochester
 CC STATE: New York
 CC COUNTRY: U.S.A.
 CC ZIP: 14603
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patented Software #1.0, Version #1.00
 CC CURRENT APPLICATION DATA:
 CC APPLICANT NAME: NIXON, HERBERT, DOWNEY & DOYLE
 CC FILING DATE: 08-SEP-1995
 CC CLASSIFICATION: 800
 CC PRIORITY APPLICATION DATA:
 CC APPLICATION NUMBER: US 07/954,847
 CC FILING DATE: 30-SEP-1992
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Goldman, Michael L.
 CC REGISTRATION NUMBER: 30,727
 CC REFERENCE/JACKET NUMBER: 19603/144
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (716) 264-1404
 CC TELEFAX: (716) 264-1600
 CC INFORMATION FOR SEQ ID NO: 18:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 69 amino acids
 CC TYPE: amino acid

CC STRANDEDNESS: 32,140
 CC IDPOLARITY: unknown
 CC MOLECULE TYPE: peptide
 CC SEQUENCE: 69 AA; 7680 MW; 21477 CH;
 Query Match: 15.5% Score 91; 108 Z; Length 269;
 Host Local Similarity: 28.6%; Prod. No. 2,090,01;
 Matches: 12; Conservative: 9; Mismatches: 20; Indels: 1; Gaps: 1;
 DB 4 FLESESPVSLTLLTLLSHSAAIMPRERIRV 41
 YY 15 YIVPLGSLVIVHAFVHDAIPPSIGQWKSKECD 52
 RESULT 15
 ID US-08 547 182-2 STANDARD: FBI: 48 AA.
 AC XXXXXX
 DE Sequence 2, Application US/0847182
 XX Patent No. 5672694
 CC GENERAL INFORMATION:
 CC APPLICANT: Campbell, Keith L.
 CC APPLICANT: Ltd, Lehigh
 CC APPLICANT: Lehigh, Frank
 CC APPLICANT: Smadar, Yashir
 CC APPLICANT: Beckman, Jacques S.
 CC APPLICANT: Brook, David
 CC APPLICANT: Toney, Donald M S.
 CC APPLICANT: Fardoux, Richard
 CC APPLICANT: Jackson, Charles E.
 CC TITLE OF INVENTION: TRANSFERRING POLYMERS FROM WITH FIVE
 CC NUMBER OF SEQUENCES: 12
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESS: Koyuk M. Little
 CC STREET: P.O. Box 999
 CC CITY: York Harbor
 CC STATE: ME
 CC COUNTRY: USA
 CC ZIP: 04911
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patented Software #1.0, Version #1.00
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US 07/947,182
 CC FILING DATE:
 CC CLASSIFICATION: 415
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Little, Koyuk M.
 CC REGISTRATION NUMBER: 40,515
 CC REFERENCE/JACKET NUMBER: 19502
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 207-363-0158
 CC TELEFAX: 207-363-0528
 CC INFORMATION FOR SEQ ID NO: 41:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 318 amino acids
 CC TYPE: amino acid
 CC MOLECULE TYPE: protein
 CC SEQUENCE: 318 AA; 34777 MW; 42144 CH;
 Query Match: 15.5% Score 91; 108 Z; Length 269;
 Host Local Similarity: 28.6%; Prod. No. 2,090,01;
 Matches: 12; Conservative: 9; Mismatches: 20; Indels: 1; Gaps: 1;

12b 65 AICVL-ILFLIIVINLIIILVIMAVIRIGPNCID-SMEFHESGLAKF 110
11:11:11 11:11:11 11:11:11 11:11:11 11:11:11
CY 22 SIYVVDIAVFIILIDAIPESTIOWKVSXKCHNOGELSDPH-OUTPOT 68

Search completed: Fri Aug 25 17:59:25 2000
Job time : 20 secs.

 M E R E N E I I (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit,
 Copyright (c) 1993 1998 University of Pittsburgh, PA
 Distribution rights by Oxford Molecular Ltd

Search: protein - protein database search, using Smith-Waterman algorithm
 Run on: Fri Aug 25 17:55:51 2000; Master time 3.96 seconds
 Tabular output not generated; 406,440 Million cell updates/sec

Title: >US-09-257-585-5
 Description: (1-68) from US09257585.ppt
 Perfect Score: 522
 Sequence: 1 MALLNEDFTLNSYLEVPI.....RRKKRKEESLPPIQVQI POL 68

Scoring table:
 Gap 11
 PAM 150

Searched: 188963 seqs, 24686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-genecol36
 1:genecol36

Statistics: Mean 27.446; Variance 111.070; scale 0.247

Prod. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Prod. No.
1	99	19.0	874	1	1:genecol36	1.29e+00
2	94	18.0	825	1	Sequence of the "gH" h	4.33e+00
3	92	17.6	1084	1	Plasmodium PASK75 open r	4.86e+00
4	92	17.6	1231	1	Sequence translated fr	4.86e+00
5	92	17.6	1277	1	Plasmodium PASK60 ST ep r	5.85e+00
6	91	17.4	527	1	H. pylori GHP 435 pro	5.85e+00
7	88	16.9	442	1	G-protein coupled rece	1.02e+01
8	88	16.9	442	1	G-protein coupled rece	1.02e+01
9	87	16.7	182	1	T cell receptor protol	1.29e+01
10	87	16.7	182	1	T cell receptor protol	1.29e+01
11	87	16.7	182	1	T cell receptor protol	1.29e+01
12	87	16.7	182	1	T cell receptor protol	1.29e+01
13	87	16.7	182	1	T cell receptor protol	1.29e+01
14	84	16.1	46	1	G-protein coupled rece	2.13e+01
15	84	16.1	46	1	G-protein coupled rece	2.13e+01
16	84	16.1	430	1	Salmonella enterica o	2.13e+01
17	84	16.1	1876	1	S. enterica enterica o	2.13e+01
18	83	15.9	69	1	S. enterica enterica o	2.13e+01
19	83	15.9	1084	1	Arabidopsis thaliana pr	2.52e+01
20	83	15.9	1876	1	S. enterica enterica o	2.52e+01
21	81	15.5	116	1	Bacillus subtilis meta	3.66e+01
22	81	15.5	211	1	Bacillus subtilis meta	3.66e+01
23	81	15.5	312	1	Synchromyces sp. beta	3.66e+01

24	81	15.5	427	1	Bacillus subtilis	3.66e+01
25	79	15.1	235	1	Bacillus subtilis	3.66e+01
26	79	15.1	318	1	Bacillus subtilis	3.66e+01
27	79	15.1	418	1	Bacillus subtilis	3.66e+01
28	79	15.1	418	1	Bacillus subtilis	3.66e+01
29	79	15.1	418	1	Bacillus subtilis	3.66e+01
30	79	15.1	418	1	Bacillus subtilis	3.66e+01
31	78	14.9	418	1	Bacillus subtilis	3.66e+01
32	78	14.9	418	1	Bacillus subtilis	3.66e+01
33	78	14.9	418	1	Bacillus subtilis	3.66e+01
34	77	14.8	259	1	Bacillus subtilis	3.66e+01
35	77	14.8	426	1	Bacillus subtilis	3.66e+01
36	77	14.8	426	1	Bacillus subtilis	3.66e+01
37	77	14.8	1294	1	Bacillus subtilis	3.66e+01
38	77	14.8	1294	1	Bacillus subtilis	3.66e+01
39	76	14.6	134	1	Bacillus subtilis	3.66e+01
40	76	14.6	366	1	Bacillus subtilis	3.66e+01
41	76	14.6	366	1	Bacillus subtilis	3.66e+01
42	76	14.6	51	1	Bacillus subtilis	3.66e+01
43	76	14.6	784	1	Bacillus subtilis	3.66e+01
44	76	14.6	784	1	Bacillus subtilis	3.66e+01
45	76	14.6	985	1	Bacillus subtilis	3.66e+01

RESULTS

1
 ID R79459 standard; Protein h1. AA.
 AC R79459;
 DI 16-JAN-1996 (first entry)
 DE Infectious laryngotracheitis virus (ILT) glycoprotein
 KW vaccine; immunosay; detection; diagnosis; identification
 KW (only) probe; antisense; gene expression.
 OS Gallus gallus
 FH Key
 FI misc_difference 656, 1000
 FT /note: "Based on no amino acids are conserved in the sequence from
 specific to general."

US5443831-A.
 PD 22-AUG-1995.
 PE 29-DEC-1991: 788124.
 PK 29-DEC-1991: US 788124.
 PR 19-NOV-1993: US 154866.
 PA (OYDE) UNIV DELAWARE.
 PI Keeler CL, Paulsen BJ.
 PR WPI: 95-302091/39.
 DR N-DS08: Q97351.
 PT Isolated infectious laryngotracheitis virus (ILT) glycoprotein (gH) for detection and for vaccine(s).
 PT Immunising pol.
 PS Claim 1: Columns 11-16: 1149: English.
 CC The infectious laryngotracheitis virus (ILT) glycoprotein (gH) can be used as a probe, for antisense, for gene expression or for production of an antibody. The polypeptide can be used for the production of an antibody (in immunosays) and in vaccines. Recombinant gH expressing the ILV gH can be used for immunisation of chickens.
 CC Sequence: 874 AA.

Query Match: 100% 99% 99% 99% 99% 99% 99% 99% 99% 99%
 Best Local Similarity: 43.0% 43.0% 43.0% 43.0% 43.0% 43.0% 43.0% 43.0% 43.0% 43.0%
 Matches: 99 99 99 99 99 99 99 99 99 99

DB 637 LARINOTRACHEITIS VIRUS (ILT) G1
 1 1 1 1 1 1 1 1 1 1
 4 LARINOTRACHEITIS VIRUS (ILT) G1
 1 1 1 1 1 1 1 1 1 1
 RESULT 2
 IF 42237 standard; Protein h1. AA

transducing domains that can be used in the membrane bound proteinaceous chimeric receptor of the invention. This sequence represents the T3 gamma receptor protein. Alternatively the transmembrane region of the chimeric receptor contains a portion of the CD7, CD5 or CD34 transmembrane domains. The extracellular portion of the receptor can also be separated from the intracellular portion by the hinge, CH2 and CH3 domains of human IgG1. The extracellular portion of the chimeric receptor contains a fragment of CD4 (amino acids 1-394 or 1-206 of the CD4 sequence) which specifically recognises and binds HIV-infected cells, but does not mediate HIV infection. The extracellular domain of the receptor is separated from the cell membrane by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices. The cells expressing the receptor are preferably T cells, B cells, neutrophils, or dendritic cells. The therapeutic cells expressing the chimeric receptor are administered to a mammal to treat HIV infection.

Query Match 16.7% Score 87 DB 1 Length 182
Best Local Similarity 45.7% Prod. No. 1,230-01

Matches 10: Conservative 12: Mismatches 5: Gaps 1: Indels 1:

DB 115 LSHFAEIVSVIAVCVFIAQDQCV 142
12 LNSYLEVPIGSIYVVDIAV FILLDAI 38

RESULT 12
ID R78670 standard; protein; 182 AA.
AC R78670:
DE 11-APR-1996 (first entry)
DE T3 gamma receptor protein.
KW Chimeric receptor; CD4; T-cell receptor; HIV; cytotoxicity;
KW human immunodeficiency virus; adoptive immunotherapy;
KW T3 gamma receptor;
OS Homo sapiens.

Key Location/Qualifiers
region 141..182
FT 400-aa "amino acids 141-182" region
FT sufficient for cytolytic signal
FT transduction"

WP9521528-A1.
FE 12-DEC-1995: 000454.
PR 14-FEB-1994: US-195395.
PR 02-AUG-1994: US-284391.
PA (GENO) GEN HOSPITAL CORP.
PI Banapour B, Kolanus W, Komuro C, Seed B.
PM: 95-292893/38.
PT Targeted cytotoxicity of HIV-infected cells - by chimeric CD4 receptor-bearing cells.
PS Example 11: Fig 17: English.
CC Intracellular and transmembrane signal transducing domains are derived from the T-cell receptor proteins CD4 and CD3 (R78670) and the T3 gamma (R78670) and from the B-cell receptor proteins IgM1 (R78671) and Ig2 (R78672). Each may be used to give extracellular domain of CD4 to obtain a chimeric receptor useful for targeted cytotoxicity of HIV-infected cells.
SQ Sequence 182 AA:

Query Match 16.7% Score 87 DB 1 Length 182
Best Local Similarity 45.7% Prod. No. 1,230-01

Matches 10: Conservative 12: Mismatches 5: Gaps 1: Indels 1:

DB 115 LSHFAEIVSVIAVCVFIAQDQCV 142
12 LNSYLEVPIGSIYVVDIAV FILLDAI 38

RESULT 13
ID M81144 standard; protein; 182 AA.
AC M81144:
DE 01-FEB-1999 (first entry)
DE T3 gamma receptor protein.

Human; zeta; eta; gamma; membrane bound chimera to (unmatured) cancer cells; autoimmune-generated cell; T3 gamma receptor; B cell receptor; Fc receptor; pathogen; bacteria; protozoan; viral.
KW Chimeric receptor; CD4; T-cell receptor; HIV; cytotoxicity;
KW human immunodeficiency virus; adoptive immunotherapy;
KW T3 gamma receptor;
OS Homo sapiens.
PM: 95-292893/38.
PT Targeted cytotoxicity of HIV-infected cells - by chimeric CD4 receptor-bearing cells.
PS Example 11: Fig 17: English.
CC Intracellular and transmembrane signal transducing domains are derived from the T-cell receptor proteins CD4 and CD3 (R78670) and the T3 gamma (R78671) and from the B-cell receptor proteins IgM1 (R78671) and Ig2 (R78672). Each may be used to give extracellular domain of CD4 to obtain a chimeric receptor useful for targeted cytotoxicity of HIV-infected cells.
SQ Sequence 182 AA:

Query Match 16.7% Score 87 DB 1 Length 182
Best Local Similarity 45.7% Prod. No. 1,230-01

Matches 10: Conservative 12: Mismatches 5: Gaps 1: Indels 1:

DB 115 LSHFAEIVSVIAVCVFIAQDQCV 142
12 LNSYLEVPIGSIYVVDIAV FILLDAI 38

RESULT 14
ID R50727 standard; protein; 182 AA.
AC R50727:
DE 11-MAY-1996 (first entry)
DE G-protein coupled receptor; G-protein coupled polypeptide;
KW G-protein coupled receptor; ligand binding assay;
KW psychotropic disorder; schizophrenia; dopamine; GPCR;
KW muscarinic acetylcholine receptor; muscarinic;
KW rhodopsin; opsin; G-protein coupled receptor;
OS Synthesis.
PM: 95-292893/38.
PT Targeted cytotoxicity of HIV-infected cells - by chimeric CD4 receptor-bearing cells.
PS Example 11: Fig 17: English.
CC Intracellular and transmembrane signal transducing domains are derived from the T-cell receptor proteins CD4 and CD3 (R78670) and the T3 gamma (R78671) and from the B-cell receptor proteins IgM1 (R78671) and Ig2 (R78672). Each may be used to give extracellular domain of CD4 to obtain a chimeric receptor useful for targeted cytotoxicity of HIV-infected cells.
SQ Sequence 182 AA:

Query Match 16.7% Score 87 DB 1 Length 182
Best Local Similarity 45.7% Prod. No. 1,230-01

Matches 10: Conservative 12: Mismatches 5: Gaps 1: Indels 1:

DB 115 LSHFAEIVSVIAVCVFIAQDQCV 142
12 LNSYLEVPIGSIYVVDIAV FILLDAI 38

RESULT 15
ID M81144 standard; protein; 182 AA.
AC M81144:
DE 01-FEB-1999 (first entry)
DE T3 gamma receptor protein.

Best Local Similarity 46.2%; Pred. No. 2.13e+01;

Matches 12; Conservative 9; Mismatches 3; Indels 2; Gaps 2;

10 19 IFNLMAPLEPGISGFYI-PIAVHLL 43

10 IF-INSYLFVPGISYVVDIAVFLL 34

RESULT 15

10 W02919 standard; peptide: 46 AA.

AP W02919;

10 04-OCT-1996 (first entry)

10 6-protein coupled receptor TM5 consensus polypeptide #9.

10 6-protein coupled receptor; ligand binding assay; transmembrane domain;

10 schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;

10 muscarinic acetylcholine; endobell; bombesin; endocrine; rhodopsin;

10 odorant; cytomodulin; serotonin; c.

10 Synthetic

10 US508384-A.

10 16-APR-1996; 941246.

10 10-SEP-1992; US-941246.

10 09-SEP-1993; US-118270.

10 (UNIV) UNIV NEW YORK STATE.

10 Murphy RB, Schuster DJ;

10 WPI; 96-208785/21.

10 New dopamine receptor peptide useful as antipsychotic agent, e.g.

10 for treating schizophrenia

10 Disclosure: Column 279-280; 184pp; English.

10 Polypeptides W02747-W02919 are based on the consensus transmembrane

10 domain III sequence from G-protein coupled receptor (GPR) proteins

10 and can be used in GPR ligand binding assays. The assays can be used

10 to identify fragments, prot. transmembrane fragments, from GPR proteins

10 (see W02657-W02655) which retain biological activity such as binding a

10 GPR ligand or modulating a GPR ligand binding to a GPR (see W02747-W02919

10 for examples of polypeptide fragments). The polypeptide fragments can be

10 used in compounds, for treating subjects suffering from a pathology related

10 to a GPR abnormality e.g. schizophrenia.

10 Sequence 46 AA;

10 Query Match 16.1%; Score 84; DB 1; Length 46;

10 Best Local Similarity 46.2%; Pred. No. 2.13e+01;

10 Matches 12; Conservative 9; Mismatches 3; Indels 2; Gaps 2;

10 19 IFNLMAPLEPGISGFYI-PIAVHLL 43

10 IF-INSYLFVPGISYVVDIAVFLL 34

Search completed: Fri Aug 25 17:56:08 2000
Job time : 17 secs.